Aalto University
School of Science
Degree Programme of Engineering Physics and Mathematics

Martti Sutinen

Adaptive Emotion Based Decision Model for a Social Robot

Master's Thesis
Helsinki, September 13, 2012

Supervisor: Professor Ahti Salo, Aalto University
Instructor: Professor David Ríos Insua, Rey Juan Carlos University

Aalto University
School of Science
Degree Programme of Engineering Physics and Mathematics

ABSTRACT OF
MASTER'S THESIS

Author: Martti Sutinen

Title of thesis:
Adaptive Emotion Based Decision Model for a Social Robot

Date: September 13, 2012 **Pages:** 14 + 74

Professorship: Systems and operations research **Code:** F200-3

Supervisor: Professor Ahti Salo
Instructor(s): Professor David Ríos Insua

This thesis introduces a computational model of emotions and decisions for a robot, which interacts meaningfully in a social context. The decision making framework is based on multi-attribute utility theory, but it contains a dynamic and adaptive emotional model which basically acts as a preference and perception manipulator.

The emotional model is based on event appraisal with discrete emotion categories. Events are assessed using dimensions of utility and probability as well as expectations. The model uses the concepts of core affect and attributed affect to create a multilevel emotion consisting of moods and emotional events. Personality traits are used to create different emotional dynamics by modifying relevant parameters. Attitudes and relationships, understood through attributed affect and classical conditioning, make the robot emotions more believable. The robot learns from user actions and makes predictions about them and environment changes according to probabilistic models.

Subjective well being and human need hierarchy are used as the basis for the preferences which the visceral state affects. The model is inspired by the computational models Cathexis, FLAME, EMA, TAME and Roboceptionist, and is an expanded version of the model used in AISoy1 robot. The framework combines extensive psychological research and requires validation.

Keywords: Emotional model, Decision making, Social robot, Event appraisal, Adaptive interaction
Language: English

Aalto-yliopisto
Perustieteiden korkeakoulu
Teknillisen fysiikan ja matematiikan koulutusohjelma

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä:	Martti Sutinen	
Työn nimi:		
Adaptiivinen Tunnepohjainen Päätösmalli Sosiaaliselle Robotille		
Päiväys:	13. syyskuuta 2012	**Sivumäärä:** 14 + 74
Professuuri:	Systeemi- ja operaatiotutkimus	**Koodi:** F200-3
Työn valvoja:	Professori Ahti Salo	
Työn ohjaaja:	Professori David Ríos Insua	

Työssä kehitetään tunteiden ja päätösten laskennallinen malli robotille, jotta se voisi käyttäytyä sosiaalisissa tilanteissa järkevästi. Päätöksentekomalli perustuu monitavoitteeseen arvoteoriaan, mutta sitä on muutettu niin, että tunnemalli säätelee mieltymyksiä sekä tulkintaa dynaamisesti ja mukautuvasti.

Tunnemalli perustuu tapahtuma-arviointiin ja diskreetteihin tunnekategorioihin. Tapahtumia arvioidaan käyttämällä hyötyä, todennäköisyyksiä sekä odotuksia ulottuvuuksina. Malli käyttää ydintunnetilaa ja liitettyä tunnetilaa luodakseen monikerroksisen tunteen, joka sisältää mielialan ja tunteikkaita tapahtumia. Persoonallisuuspiirteillä saadaan aikaan monipuolisia tunnedynamiikkoja säätämällä asianmukaisia parametreja. Asenteet ja suhteet tekevät robotista uskottavamman, ja ne ymmärretään liitetyn tunnetilan sekä klassisen ehdollistamisen avulla. Robotti oppii käyttäjän teoista ja tekee niistä sekä ympäristön kehityksestä ennusteita todennäköisyyspohjaisilla malleilla.

Subjektiivinen hyvinvointi ja ihmisen tarvehierarkia antavat mieltymysten painotukselle pohjan, jota robotin sisäinen tila muokkaa. Laskennalliset mallit Cathexis, FLAME, EMA, TAME and Roboceptionist inspiroivat mallia, joka on laajennettu versio AISoy1 robotin mallista. Kehys yhdistää laajaa psykologista tutkimusta ja vaatii testausta.

Avainsanat:	Tunnemalli, Päätöksenteko, Sosiaalinen robotti, Tapahtuma-arviointi, Mukautuva vuorovaikutus
Kieli:	Englanti

Acknowledgements

This thesis would not have been completed without the help of many wonderful, loving and supporting characters. I want to start by thanking my mother Päivi, who has been watching after me since I was born. I am sure that it has not always been easy. My father Erkki deserves credit for pushing me forward academically and raising my ambitions. He suggested that I should try finishing my studies in four years. My siblings Anton, Laura and Tuuli endured my occasional stressful behaviour admirably. I had various conversations with my friends Mikko, Ville, Anton and others, and my philosophy teacher Esa, who were all encouraging in their own ways. I had the priviledge to get Ahti as my supervisor, after he suggested contacting my exceptionally talented instructor, David. I was lucky that they both were extremely patient with me while I asked a lot of questions and offered original ideas to overcome the bureaucracy related to writing the thesis during exchange studies.

Before leaving to Madrid, an old friend, Andrés, connected me with his family. It was extremely valuable to have a second home always open for me. The tasty breakfasts and dinners gave me a chance to get a break from work and practice my Spanish with Marisa and Paco. My colleagues Pablo, Alberto and Javier made sure that I got a lively impression of the university and the whole city. They shared their time, knowledge, lunches, table tennis games and parties with me to help me find new friends and learn the culture. I could laugh with them when I was too Finnish to do it on my own.

I spent a lot of time with another thesis writer in the group, Ivan, and his girlfriend Kathy. They were living perhaps too close to me because I ended up buying a leg of a pig together with them. It was delicious, just as everything else we prepared. My short and exciting visits to the neighbouring countries were accompanied by Samu, Olli, Tuomo, Roberto, Marco and Sean.

David did an extraordinary effort in guiding me through the thesis. He was flexible and innovative in adjusting schedules and gathering material. I was able to finish my courses in time because he and Kimmo agreed to organize

an important long distance exam. Ahti was also understanding about the distinctive writing process.

One person occupied my thoughts and emotions more than any thesis could. Essi was brave enough to decide with me to keep our relationship alive and make the most of it during my exchange studies. Our calls, letters, visits and connection never left me cold. There are no words to describe the sensations I have experienced during our time together, and I will not attempt to convert them into mathematics.

Helsinki, September 13, 2012

Martti Sutinen

Contents

Abbreviations and Acronyms VIII

Notations VIII

List of tables XII

List of figures XII

1. Introduction 1

1.1. Emotions and decisions 1

1.2. Affective robots and models 2

1.2.1. Decision Affect Theory 3

1.2.2. Cathexis 4

1.2.3. FLAME 5

1.2.4. EMA 7

1.2.5. TAME 8

1.2.6. Decision Field Theory 10

1.2.7. Roboceptionist 11

1.2.8. AISoy1 12

1.3. Research objectives 13

2. Theories of emotions 15

2.1. Somatic basis 15

2.1.1. Brain structure 15

2.1.2. Somatic marker hypothesis 18
2.2. Psychological models 19
2.2.1. Dimensions of emotions 20
2.2.2. Core affect, emotional events and moods 24
2.2.3. Conditioning and attitudes 25
2.2.4. Expectations and alternatives 27
2.2.5. Personality traits for diverse dynamics 27
2.2.6. Human needs and subjective well-being 31
2.3. Emotions in economics 32
2.4. Connection with game theory 34
2.5. Effect of emotions on decisions 36

3. Modeling affective decision making **39**
3.1. Decision model 39
3.1.1. Framework and sets 39
3.1.2. Forecasting models 40
3.1.3. Action selection 42
3.1.4. Objectives and utilities 43
3.2. Emotional model 43
3.2.1. Mood 44
3.2.2. Emotions 47
3.2.3. Attitude 52
3.2.4. Fear and pain 52
3.2.5. Heuristic action evaluation 53
3.2.6. Effects on the Decision Model 54
3.3. Robot features 56
3.4. Further research 56

4. Conclusions **59**

Bibliography **61**

A. Vytal's review of neurological discrete emotion studies 71

B. Ortony's list of basic emotion studies 72

C. Roseman's event appraisal experiment results 73

D. The model usage diagram 74

Abbreviations and Acronyms

3D	Three dimensional
ALE	Activation likelihood estimation
ANS	Autonomic nervous system
BG-DA	Basal ganglia-dopamine system
CS	Conditioned stimuli
DAT	Decision Affect Theory
DLPFC	Dorsolateral prefrontal cortex
EA	Event appraisal
EEG	Electroencephalography
GDP	Gross Domestic Product
HCI	Human-computer interaction
LS	Libmic system
NLP	Natural language processing
OFC	Orbitofrontal cortex
QOL	Quality of Life
SMH	Somatic marker hypothesis
SWB	Subjective Well-Being
US	Unconditioned stimuli
VM	Ventromedial prefrontal cortex

Notations

a_{ec}^{em}	Affine function parameter for emotional content of emotion em
a_{ec-}^{em}	Affine function parameter for expected emotional content of emotion em
$a_1, \ldots, a_m$	Agent actions
ah_t	Momentary analiticity
ap_t^i	Activation point parameter of emotion i
$\mathbf{at}_t$	Attitude vector
$\mathbf{at}_t(B_u)$	User-specific attitude vector
$\mathbf{at}_t(def)$	Default attitude vector
$\underline{at_{md}}$	Threshold of attitude for influence on mood
A	Agent
A_{mdmb}	Ratio strength parameter for mood and mood baseline
A_{mdvs}	Ratio strength parameter for mood and visceral core affect
A_{pn}	Pain normalization factor
A_{ps}	Processing strategy normalization factor
A_{vsmb}	Ratio strength parameter for visceral core affect and mood baseline
$\mathcal{A}$	Action set of the agent
α	Q-learning rate
b_{ec}^{em}	Affine function parameter for emotional content of emotion em
b_{ec-}^{em}	Affine function parameter for expected emotional content of emotion em
$b_1, \ldots, b_n$	User actions
$B_1, \ldots, B_r$	Users
$\mathcal{B}_u$	Set of users

$\mathbf{ca}_t^{vs}$	Visceral core affect vector
cf	Consistency factor
$df(x)$	Disappointment function
dp_t	Disappointment
$D_{x,y}$	Euclidean 2-norm distance between x and y
δ_p	Decision weight function attractiveness parameter
δ_{pd}	Pain decay
δ_t^{mb}	Mood baseline weight for current mood
δ_t^{md}	Previous mood weight for current mood
δ_t^{vs}	Visceral core affect weight for current mood
$\mathbf{e}_t$	Environment state vector
ec_t^{em}	Emotional content of emotion em
ec_{t-}^{em}	Expected emotional content of emotion em for time t
$\mathbf{em}_t$	Emotion intensity vector
em_t^i	Intensity of emotion i
E	Environment
$\mathcal{E}$	Environment set
$fam_t^{B_u}$	Familiarity measure for user B_u
$g(\mathbf{s}_t)$	(Probabilistic) user action evaluation function
gr^i	Growth parameter of emotion i
γ	Q-learning discount factor
γ_p	Decision weight function discriminality parameter
$h(\mathbf{s}_t)$	User interaction probability function
I_{dmax}	Maximum intensity difference
kn	Negative difference exponent parameter for $df(x)$
kp	Positive difference exponent parameter for $df(x)$
$\mathbf{mb}$	Mood baseline vector
$\mathbf{md}_t$	Mood vector
$\mathbf{M}_{ca}(i)$	Core affect mapping of visceral factor i
$M_{pe}(i,j)$	Mapping of the influence of personality trait j on emotion i
N_{em}	Amount of emotions
$N_{M_{pe}\neq 0}$	Amount of nonzero personality-emotion mappings for emotion i

$\hat{p}_t$	Scenario probability estimate
per	Personality vector
$\mathbf{pc}_t$	Physical condition vector
pv^i	Peak value of emotion i
$P(a_t)$	Probability of choosing action a_t
$\psi(a_t, \ldots, a_{t+r})$	Expected utility function of a strategy
$Q(st_t, a_t)$	Q-value for state st_t and action a_t
R_t	Reward in Q-learning
ra_t	Risk attitude
$s_1, \ldots, s_n$	Sensor readings
$\mathbf{s}_t$	Sensor reading vector
$\sigma^+_{\hat{U}_t}$	Upper standard deviation of utility
$\sigma^-_{\hat{U}_t}$	Lower standard deviation of utility
t	Time (period)
T_{mdu}	Mood model update time
$\hat{u}_t$	Scenario utility estimate
$\hat{U}^{sadness}_{high}$	Threshold for experiencing sadness
$\hat{U}_t$	Scenario expected utility estimate
$\mathbf{vs}_t$	Visceral state vector
$\underline{vs^i_{md}}$	Threshold of visceral factor i for influence on mood
$w(p)$	Decision weight function for probability
w_{at}	Core affect weight of attitude
$w_t^{at,new}$	New attitude weight
$w_t^{at,old}$	Old attitude weight
$w_t^{ob,i}$	Normalized weight of objective i
w_{vs}	Core affect weight of visceral factor vs
W_t^i	Weight of objective i

List of Tables

2.1. Different models of discrete emotions 22
2.2. Correlates of personality factors (McCrae & P. T. Costa 1997) . . 28
2.3. The utilities in prisoner's dilemma, utility pairs a,b representing the utilities of player 1 (P1) and player 2 (P2), respectively . 35
2.4. The effects of discrete emotions on risk perception 38

3.1. The action sets for the robot and the users 39
3.2. Core Affect mappings 45
3.3. Personality effects on emotions $M_{pe}(i, j)$ (Moshkina 2011) 50
3.4. Visceral state categories 54

List of Figures

1.1. The framework of Cathexis (Velásquez 1998) 4
1.2. The framework of FLAME (El-Nasr, Yen & Ioerger 2000) 5
1.3. Emotion intensities in FLAME (El-Nasr et al. 2000) 6
1.4. The computational cognitive-motivational-emotional system of EMA (Gratch & Marsella 2004) 7
1.5. Emotion intensities in EMA (Gratch & Marsella 2004) 8
1.6. The framework of TAME (Moshkina 2006) 9
1.7. Influences of personality and mood on emotion generation in TAME (Moshkina 2006) 10
1.8. The decision process in ARA (Esteban 2012) 12

2.1. Brain areas related to decision making, bold text indicating areas related to emotional processing and italics to cognitive processing (Cohen 2005) 16
2.2. Neural circuits of fear conditioning (LeDoux 1995) 17
2.3. Differences between choices of gambling decks of normal people and VM patients (Bechara, Damasio & Damasio 2000) 19
2.4. Russell's two-dimensional model with discrete emotions plotted (Russell & Barrett 1999) 21
2.5. Roseman's event appraisal model (Roseman, Spindel & Jose 1990) 24
2.6. Comparison of distinctions between emotion and mood (Beedie, Terry & Lane 2005) 26
2.7. Markon's structural model of personalities, N=Neuroticism, A=Agreeableness, C=Concentiousness, E=Extraversion, O=Openness (Markon, Krueger & Watson 2005) 28

2.8. Correlations between personality and mean level and standard deviation of emotions (Eid 1999) 29
2.9. The circumplex structure of personality by Heller (1993) (Schmidtke & Heller 2004) 30
2.10. DynAffect model correlations between dispositions and parameters with p-values in parentheses (Kuppens, Oravecz & Tuerlinckx 2010) 30
2.11. Probability weighting function inspired by Tversky & Kahneman (1992) (Loewenstein & Lerner 2003) 37

3.1. The objectives pyramid (Esteban 2012) 43
3.2. An example of a core affect calculation with intensities $\mathbf{vs}_t = [0.1, 0.1, 0.3, 1, 0.5, 1]$, weights $\mathbf{w}_{vs} = [2, 1, 1, 2, 0.5, 0.5]$ and attitude $\mathbf{at}_t = (0.5, -0.2)$ with a weight $w_{at} = 0.5$ 46
3.3. An example of a mood change 47
3.4. Emotion intensity function examples 51
3.5. The decision weight function 54
3.6. AISoy1 Robot (AISoy Robotics S.L. 2012) 56

A.1. Brain activation likelihood maps for discrete emotions (Vytal & Hamann 2010) 71

B.1. Different studies of basic emotions (Ortony & Turner 1990) . . . 72

C.1. Mean appraisal ratings and significance tests for appraisal dimensions (Roseman et al. 1990) 73

D.1. The model usage process with solid arrows indicating progress and dashed arrows indicating data flow 74

Chapter 1

Introduction

1.1. Emotions and decisions

Traditionally, decision theory has acknowledged emotions as a biasing factor in decision making and sought to remove it from the decision making process. When an animal is threatened by a predator, it can be immobilized by fear even if an escape would be more useful (Janis & Leventhal 1967). Earthquake insurances are bought more after earthquakes while the probability of a new accident stays constant (Palm, Hodgson, Blanchard & Lyons 1990). It seems then that emotions should be suppressed in decision making, because they modify preferences and judgements.

However, emotions often appear purposeful when they are not extreme. When seeing a bear raises fear, a flight mechanism can help in reacting quickly instead of assessing probabilities and utilities between climbing a tree, running or fighting (LeDoux 1995). Indeed, new research in neurology shows that it is very difficult for humans to act effectively without emotional capabilities (Damasio 1994). As an example, a patient called Elliot, a man with a family and a job, suffered damage in his ventromedial prefrontal cortex in a surgery. He scored normal in neurological and intelligence tests but he could not make everyday decisions smoothly anymore. He lost his wife and his job, and had to spend half an hour to conduct cost-benefit analysis in choosing between two dates. Railway worker Phineas Cage suffered damage in the same part of the brain and his life failed afterwards as well (Damasio & Grabowski 1994). This brain area is responsible for linking events with emotional states (Bechara et al. 2000).

Not only are emotions useful for decision making but they are such an in-

tegral part of it that they cannot be ignored when realistic decision models are constructed. In the ultimatum game, two players split a given amount of money. The first player chooses how much he keeps and how much is offered to the second player. The second player can accept the offer, or reject it and neither player gets anything. In theory the first player should offer as little as possible and the second player should accept because a small amount is better than nothing. But unfair offers often cause anger and are rejected, which can be seen as an activation in brain areas related to emotion (Sanfey, Rilling, Aronson, Nystrom & Cohen 2003).

In politics, voters may shift their preferences depending on their mood. Anxiety and fear promote attention to political argumentation as focus is on problem solving and survival. New information is searched and candidates are challenged. On the other hand, enthusiasm and joy cause voters to be more habitual, and emotional associations with candidate party and background are more important (Marcus & Mackuen 1993). Even very serious and influential decisions can be based on emotions. It has been suggested that part of president Bush's motivation to start the war in Iraq was revenge for Hussein's earlier attempt to assassinate his father (McDermott 2004).

Emotions are not perfectly defined in the sense that there is still debate on what emotions are (Russell 2003). Especially important is whether there are some basic emotions (Ekman & Friesen 1971, Ortony & Turner 1990, Levenson, Ekman, Heider & Friesen 1992, Izard 1992, Panksepp 1992, Turner & Ortony 1992), how emotions evolve (Zajonc 1984, Lazarus 1984, Frijda, Kuipers & ter Schure 1989, Roseman et al. 1990), what the dimensions of emotions are (Russell & Barrett 1999, Larsen & McGraw 2001) and what their effects on decisions are (Leone, Perugini & Bagozzi 2005, Forgas 1995). But it is clear that emotions affect decisions in several ways, from weighing criteria and evaluating alternatives (Lerner & Keltner 2000) to mood congruency of memories (Rusting & DeHart 2000) affecting predictions.

1.2. Affective robots and models

The new research in psychology and neuroscience has inspired emotion, which is often called affect (Russell & Barrett 1999) when it is coupled with similar concepts such as mood (Beedie et al. 2005), to be taken as part of computational models of intelligent agents. Affective computing combines the research to produce machines which are able to recognize, model and communicate emotions to enhance human computer interaction (HCI) and aid

related research in suprising ways (Picard 1997, Picard 2003). For example, Decision Affect Theory (Mellers, Schwatz, Ho & Ritov 1997) provided empirical evidence of the effect of expectations on emotion generation, which could be used to support motive-consistency as a dimension in Roseman's event appraisal theory (Roseman et al. 1990). Extensive models such as TAME (Moshkina 2006) create demand for research connecting different theories like personality traits and their effects on emotions. Several computational models for emotions have been created, some with physical implementation, which we briefly describe.

1.2.1. Decision Affect Theory

Decision Affect Theory (DAT) translates risk, or probability, and expectations for events to a one-dimensional feeling which might be compared to the valence dimension of Russell (2003). It is motivated by regret and dissappointment theories, and tries to capture the effect of other alternatives on experienced benefit of an outcome. Equation 1.1 shows the formula for an emotional response to an outcome. The theory was experimented with different obtained and unobtained outcomes and probabilities. Mellers et al. (1997) achieved an excellent fit with less than 1% residual variance. The model is

$$R_a = c \cdot \left[u_a + \sum_{b \neq a} g(u_a - u_b) \cdot s_b \right] + d, \tag{1.1}$$

where R_a is the emotional response to outcome a, $c > 0$ and d are coefficients in the judgement function, u_a and u_b are utilities of the obtained and unobtained outcomes, s_b is the subjective probability of event b and $g()$ is a dissappointment function. A larger emotional response implies the superior emotional utility of an outcome.

The effect of emotional responses of outcomes in gambles on choices between gambles was also examined. Subjective expected emotion was calculated as a sum of expected responses for outcomes $\sum s_i R_i$. The correlation between binary choices and expected emotion predictions of the gambles was 0.89 suggesting that the emotional content of alternatives was a major factor in the decisions. This is similar to the idea of the somatic marker hypothesis (Bechara et al. 2000). DAT shows that expectations and risks affect emotions but does not differentiate between discrete emotions. (Mellers et al. 1997)

1.2.2. Cathexis

Velásquez' Cathexis model uses different sources of releasers for emotion generation and for other systems. The releasers are either natural, e.g. signals resulting from various sensors or chemical reactions in brain, or learned releasers, e.g. conditioning sound of hitting with pain and fear. The activation of basic systems is a nonlinear function, such as a standard ramp function, of weighted sum of relevant releasers for the system. Also the states of other systems can be input for activation functions. For example, the hunger drive system could cause the distress emotional system to activate. Emotional systems also include a decay function in the activation function variable. Releasers are divided in four categories: neural, sensorimotor, motivational and cognitive, inspired by Izard (Izard 1993). The activation function variable (Velásquez 1998) can be seen in the equation

$$A_i(t) = f\left(\Psi_i(A_i(t-1)) + \sum_k R_{ki}W_{ki} + \sum_l \mu_{li}A_l(t)\right), \tag{1.2}$$

where $A_i(t)$ is the activation of emotional system i at time t, $\Psi_i()$ is temporal decay function of emotion i, R_{ki} is the value of releaser k and W_{ki} is the corresponding weight for the emotion, μ_{li} is the excitatory (positive) or inhibitory (negative) strength of emotional system l, and $f()$ is a limiting function such as a standard ramp or a logistic sigmoid function.

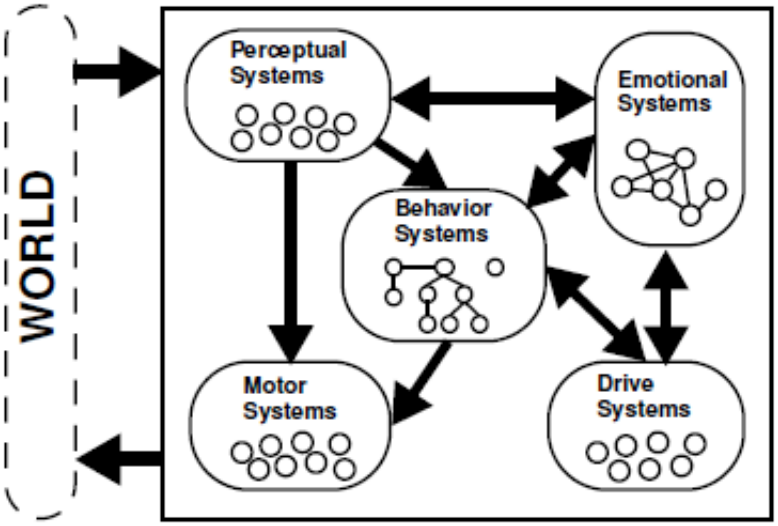


Figure 1.1: The framework of Cathexis (Velásquez 1998)

Behavioural systems are activated with a similar function of releasers and other system activity but without decay. Emotional conditioning happens through learned releasers. Weights for releasers that are simultaneous with an activated system are increased. The model framework is shown in Figure 1.1. A downside is that the model does not learn to expect future events. Our model does not include neural or sensorimotor processes, but motivational states are adapted.

1.2.3. FLAME

FLAME is a model which uses both event appraisal, formulated through fuzzy rules, and motivational states, such as pain, for emotion generation. Emotions are based on evaluating the effect of events on goals and they are used in selecting behaviour. Event appraisal is based on Roseman's (Roseman et al. 1990) and Ortony's (Ortony, Clore & Collins 1988) theories, and Bolles' fear-pain model (Bolles & Fanselow 1980). The architecture can be seen in Figure 1.2. (El-Nasr et al. 2000)

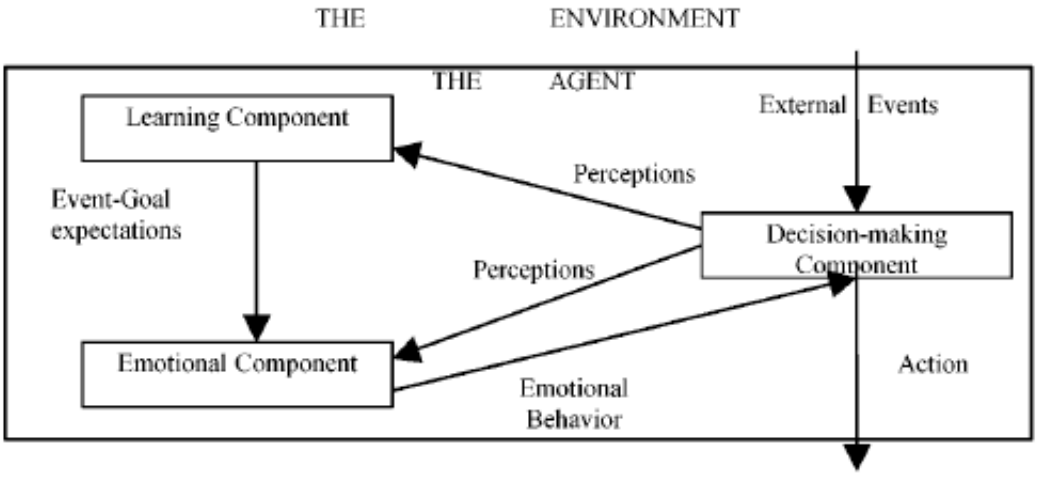


Figure 1.2: The framework of FLAME (El-Nasr et al. 2000)

The impact of an event on a goal and the importance of a goal, both represented as fuzzy sets, are used to calculate the fuzzy desirability with a rule database. Desirability is finally defuzzified with Mamdani's centroid defuzzification (Mamdani & Assilian 1975). Desirability and learned expectation, or probability, of an event elicit emotions according to Ortony's model. Some parameters are taken from Price's model (Price, Barrell & Barrell 1985). Figure 1.3 shows the calculations of corresponding emotion intensities. Altogether

fourteen emotions are modeled. In addition to desirability, the agent's standards, or moral, as well as learned attitude towards objects are used for some of them. Emotions decay in time as a product of a decay factor and previous intensity. Motivational states can also affect emotion e.g. when pain inhibits fear. Intense motivational states interrupt cognitive processes to choose relevant goals. Mood is calculated as a moving sum of positive and negative emotion intensities and is either positive or negative. Moods are used to inhibit opposite emotions. (El-Nasr et al. 2000)

Emotion	**Formula for intensity**
Joy	$Joy = (1.7 \times expectation^{0.5}) + (-0.7 \times desirability)$
Sadness	$Sad = (2 \times expectation^{2}) - desirability$
Disappointment	$Disappointment = Hope \times desirability$
Relief	$Relief = Fear \times desirability$
Hope	$Hope = (1.7 \times expectation^{0.5}) + (-0.7 \times desirability)$
Fear	$Fear = (2 \times expectation^{2}) - desirability$

Figure 1.3: Emotion intensities in FLAME (El-Nasr et al. 2000)

Behaviour is selected with fuzzy rules that consist of emotional states, events and causes of events. Selection is made based on the highest score. The agent learns in several ways. It uses average intensity of events relating to an object to simulate classical conditioning. Kaelbling's Q-learning is used to relate state - action pairs to expected reward values (Kaelbling, Littman & Moore 1996). It is modified to include probability distributions of future states because of the uncertainty regarding the user's actions and respective rewards. Mood is used to modify the belief about positive and negative rewards. A user model keeps in memory the frequency of user actions given two past actions, i.e. updates conditional probabilities of action chains of three steps. A user action leads to a new state that is used in the Q-learning valuation. An alternative approach using an average score of feedback value to actions can be used for selecting behaviour. The model is implemented in a virtual environment. A survey with 21 subjects revealed that learning is vital for making the virtual agent believable. (El-Nasr et al. 2000)

FLAME is an extensive model and has some very relevant components. The logic of using desirability, or utility, and expectations, or probability, for event appraisal is adapted to our model. Also a split user model is used for forecasting user actions, taking into account user action history and reactions. Attitudes are used to create emotional relationships with users. However, a decision theoretic approach is used for selecting actions and fuzzy logic is not

used, as it complicates forecasting and the survey did not reveal clear benefits compared to real numbers.

1.2.4. EMA

Gratch introduces a domain independent model, EMA, for emotions. It is based on event appraisal based emotion generation but it uses the idea of coping strategies to react to the environment. The model focuses on a cognitive emotion system and leaves out motivational states and reactions. Coping strategies aim at changing the environment through actions (problem-focused coping) or the interpretation of events (emotion-focused coping) to better satisfy the subject in relation to goals. The used emotional system is shown in Figure 1.4 and is adapted from Smith and Lazarus' theory (Smith & Lazarus 1990). (Gratch & Marsella 2004)

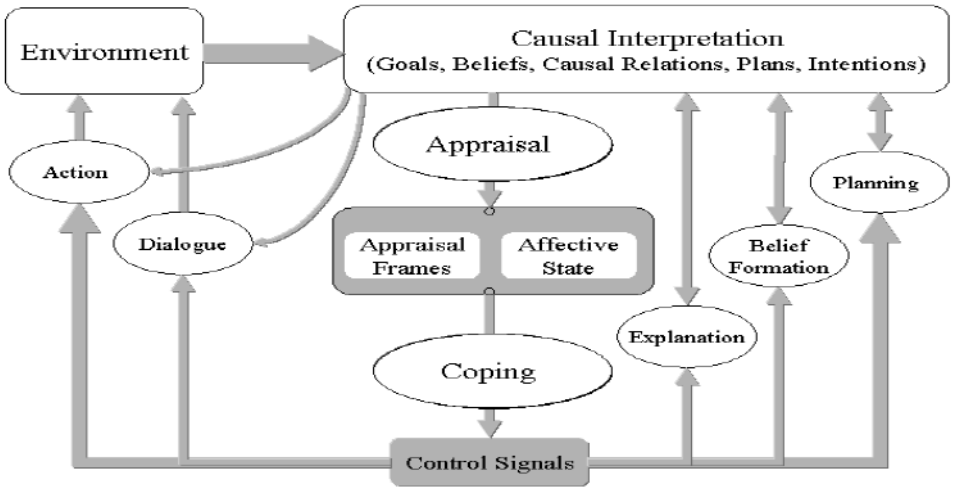


Figure 1.4: The computational cognitive-motivational-emotional system of EMA (Gratch & Marsella 2004)

The algorithm interprets first world events in terms of some appraisal dimensions. Then, emotions are calculated based on a simplified version of Elliot's Affective Reasoner (Figure 1.5) and a coping strategy is chosen (Elliott 1992). Examples of coping strategies are active coping: taking active steps to try to remove or circumvent the stressor, and seeking social support for emotional reasons: getting moral support, sympathy, or understanding. (Gratch & Marsella 2004)

EMA is implemented in a virtual reality training environment called Mission

Appraisal Configuration	Emotion	Intensity
Desirability(p) > 0, Likelihood(p) < 1.0	Hope	Desirability(p)×Likelihood(p)
Desirability(p) > 0, Likelihood(p) = 1.0	Joy	Desirability(p)×Likelihood(p)
Desirability(p) < 0, Likelihood(p) < 1.0	Fear	|Desirability(p)×Likelihood(p)|
Desirability(p) < 0, Likelihood(p) = 1.0	Distress	|Desirability(p)×Likelihood(p)|
Desirability(p) < 0, causal attribution(q)=blameworthy	Anger	|Desirability(p)×Likelihood(p)|
Desirability(q ≠ p) < 0, causal attribution(p)=blameworthy, causal agent = p	Guilt	|Desirability(q)×Likelihood(p)|

Figure 1.5: Emotion intensities in EMA (Gratch & Marsella 2004)

Rehearsal Exercise (MRE) system and is incorporated to the Steve -agent architecture (Rickel & Johnson 1999). Multiple characters in the environment are controlled by intelligent agents using EMA. These characters play the roles of locals, friendly and hostile forces, and other mission team members. MRE is a broad 3D system which also covers natural language processing (NLP) and dialogue management as well as environment perception and gestures including facial expressions. However, EMA is very different from our model because we do not use the concept of coping and behaviour is controlled by expected utility maximization. (Gratch & Marsella 2004)

1.2.5. TAME

The TAME - Traits (Personality), Attitudes, Moods and Emotions - model is a very extensive emotional model. It uses the Big Five (Openness, Concientiousness, Extraversion, Agreeableness and Neuroticism) personality trait theory that has been shown to be consistent across cultures (McCrae & P. T. Costa 1997). Traits influence behavioural parameters via functional mappings, e.g. the degree of neuroticism would be proportional to obstacle avoidance parameter and inversely proportional to wander parameter. Also, they have an effect on emotion generation like mood. Attitudes are modeled as an object dependent real number which is affected by mood, and they guide behaviour away from aversive objects. Emotions are discrete. The framework of TAME is shown in Figure 1.6. (Moshkina 2006)

For emotion generation, TAME adapts five properties from Picard (1997): activation (as in Cathexis), saturation, response decay, limited linearity, and personality and mood influences. Interestingly, personality affects the activation point, peak response and rise time to peak for an emotional stimulus. Mood varies the threshold of experiencing emotion. The equation for base emotion is shown in Equation 1.3 and the influences of personality and mood

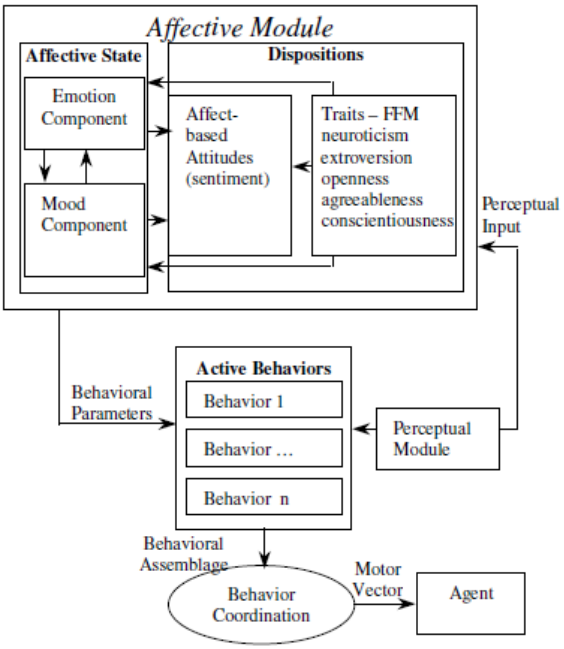


Figure 1.6: The framework of TAME (Moshkina 2006)

on it can be seen in Figure 1.7. The current emotion is a weighted average of decayed base emotion and previous emotion. Moshkina (2006) models the base emotion with

$$E_{i,base} = \begin{cases} e^{(s_i - a_i)/d_i} - e^{-a_i/d_i} & \text{if } a_i \leq s_i < (a_i + b_i)/2 \\ g_i - e^{-(s_i - b_i)/d_i} & \text{if } s_i \geq (a_i + b_i)/2 \end{cases}, \tag{1.3}$$

where $b_i = d_i \ln(g_i + e^{-a_i/d_i}) + a_i$, $E_{i,base}$ is the base value of emotion i, $s_i > 0$ is the emotional signal strength, $a_i > 0$ controls the activation point, $d_i > 0$ controls the maximum slope, $b_i > 0$ controls the point where growth rate is reversed and $g_i > 0$ is the peak value.

Moods evolve though environmental, e.g. light, battery level, and situational, e.g. emotional episodes, variables and are longer lasting than emotions. Mood is two-dimensional, consisting of positive and negative mood variables. The effect of environmental variables is propotional to their distance from neutral level. Emotions contribute according to their valence and intensity. Attitudes towards object develop through perceived attributes and are directed according to mood. TAME offers an interesting implementation for multilevel affect. Our model shares the Big Five personality trait theory and mood evolution but mood is represented in pleasure and activation dimensions. We

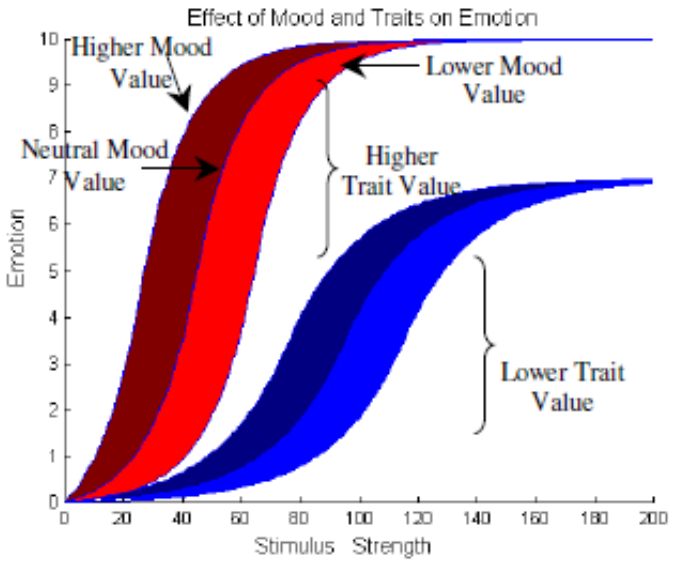


Figure 1.7: Influences of personality and mood on emotion generation in TAME (Moshkina 2006)

understand attitudes through emotional conditioning and they are discrete emotional tags on affetive objects, or people, that affect and are affected by current emotions. (Moshkina 2006)

1.2.6. Decision Field Theory

Busemeyer provides an integrated approach, Decision Field Theory (DFT), for emotional decision making that is closer to decision theory than the previous models. The model consists of a dynamic stochastic preference model which activates one alternative when a threshold value is reached. The weights w_{ij} for alternative i and consequence j are used with values m_j in the current utility function. (Busemeyer, Dimperio & Jessup 2007)

A stationary stochastic process $W_{ij}(t)$, $E(W_{ij}) = w_{ij}$, determines momentary weights. The utility is then $U_i(t) = \sum W_{ij}(t)m_j$ with mean $u_i(t) = \sum w_{ij}(t)m_j$. Consequence values m_j are determined by an affective mechanism. To calculate preference states, valence is calculated as the difference between an alternative's utility and the average utility $v_i(t) = U_i(t) - \overline{U}(t)$. The preference can be then modified with a linear dynamic stochastic difference equation $P_i(t+h) = \sum s_{ij}P_j(t) + v_i(t+h)$, where feedback coefficients s_{ij} are used to produce a competitive system. (Busemeyer et al. 2007)

To determine values m_j, the concept of needs n_k, including emotions, is used. Unsatisfied needs accumulate over time and their weight for m_j increases. The value q_{jk} for consequence j regarding attribute k is used to form the weigted consequence value $m_j = \sum n_k q_{jk}$. Needs develop through linear dynamic equations $n_k(t+h) = L_k n_k(t) + [g_k - a_k(t+h)]$, where L_k is a feedback constant, g_k is a goal level and $a_k(t+h)$ is the current level related to need k. (Busemeyer et al. 2007)

DFT is an interesting formulation but does not explain the evolution of emotion in detail. However, it does use normalized utility in calculating valence which is supported by suggestions that the value of an alternative depends on other alternatives and that framing has a significant effect on choice (Tversky & Kahneman 1992). In our model, decisions are made in a reactive manner in contrast to the threshold approach used by DFT.

1.2.7. Roboceptionist

Roboceptionist is a social robot that builds long-term relationships with people in that it assists them in basic things such as finding places at a university. It has emotions, moods and attitudes. Kirby, Forlizzi & Simmons (2010) follow Ekman & Friesen (1971) in choosing subset of joy, sadness, disgust and anger as basic emotions. Both exprienced past events and anticipated future events affect the one-dimensional mood value according to their strength and distance along a sigmoid curve. Daily base mood is then a sum of different events' contributions to it. Displayed mood changes smoothly after events, following a logistic function. (Kirby et al. 2010)

Emotions are triggered by communication and emotional content is interpreted using hand-coded rules. Only one emotion is shown at a time and it lasts a short time, so normally mood is displayed. Mood affects emotional reactions. The experienced emotion depends on the intensity and valence of the emotion, and mood. Attitude is one-dimensional and long-lasting. The evolution depends on a familiarity measure that combines the time since the previous interaction and total interaction time. After an interaction, the attitude is updated proportional to the mood change during interaction and the additive inverse of familiarity. Using familiarity in attitude formation is a good idea and it is adapted from this model. (Kirby et al. 2010)

1.2.8. AISoy1

This model, based on Adversarial Risk Analysis (ARA) (Ríos Insua, Ríos & Banks 2009) framework, is a decision theoretic approach to affective decision making in a multiactor environment. The model is implemented in the AISoy1 robot. It uses forecasting models to make predictions about future user actions and environment states. The models are updated as new information is used for learning about users. Forecasts and learning make it possible for the agent to assess alternatives in a strategic manner and take into account dynamics in an interactive environment. The decision process is shown in Figure 1.8. (Esteban 2012)

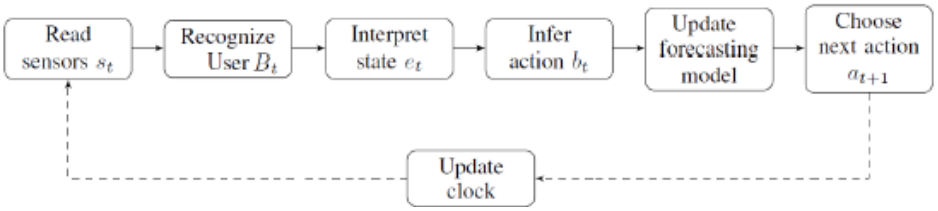


Figure 1.8: The decision process in ARA (Esteban 2012)

The framework consists of an agent A that is controlled by the model and human users $B_1, \ldots, B_r \in \mathcal{B}_u$ who interact with the agent in an environment E. A has a finite action set $\mathcal{A} = \{a_1, \ldots, a_m\}$ and users $B_u's$ have a different interpreted action set $\mathcal{B} = \{b_1, \ldots, b_n\}$ including a *do nothing* action. The environment E has dynamic states within a set $\mathcal{E}$. (Esteban 2012)

A reads external environment with sensors q at times t and the sensor reading vector is $s_t = (s_t^1, \ldots, s_t^q)$. Using the readings, A infers the environment state, current user and previous action with probabilistic functions. Future states are then evaluated with conditional probabilities calculated with those functions. The consequences of states and actions are input for the utility function which is used in action selection. Dynamic programming solves the maximum expected utility. (Esteban 2012)

Emotions in the model are discrete and based on Gratch, Marsella, Wang & Stankovic (2009) (Rázuri, Esteban & Ríos Insua 2011). They evolve based on events, i.e. robot and user actions, and are mixed in the sense that multiple emotions can occur simultaneously. The emotional state is updated with a probabilistic function that considers previous emotions, event utility and distance to expectation. Utility is in turn affected by emotions that determine criteria weights. The framework is the basis of our decision making model and the presented emotional model is integrated into it. Formulas are presented

in detail in Chapter 3. (Esteban 2012)

1.3. Research objectives

The thesis contributes to research of emotions in an attempt to provide a suitable and expandable model for human-computer interaction in social robotics. The proposed emotional model is combined with the AISoy1 decision theoretic framework that explains behaviour based on optimizing expected utility in a multiactor environment. The thesis is inspired by theories and experimental results from psychology, neurology, economics, game theory and decision theory as well as existing computational models. The emotional model will include the following features and relevant theory will be reviewed:

- multilevel affect: mood and discrete emotions
- versatile emotion dynamics: event appraisal, mood congruency, affective objects and motivational states
- relationships through conditioning and familiarity-dependent attitude formation
- personalities that affect emotion generation and behaviour
- effect of emotions on perceiving utility

The decision model includes adaptive forecasting models that are used to make predictions about future user actions and environment state. To simplify the robot operation and to reduce computational time allowing real-time interaction, the following features are not yet included:

- learning about the utility of the user based on interaction
- learning new actions and their consequences
- understanding group interaction and active participants
- goal selection and relevant task management for goals that cannot be completed with a single action

The thesis is organized in the following structure. In this Chapter, we have introduced different implementations of emotional models and affectionate robots, as well as defined the research objectives. In Chapter 2, the basis for emotional models is discussed, and the relevant neurological and psychological research is presented. Using the theories, implications for mathematical models of emotional decision making are reviewed. Chapter 3 contains the mathematical structure of the developed emotional decision making model, divided in the decision making and the emotional system components, with some recommendations for further research. Finally, the introduced model is summarized in Chapter 4 and conclusions are drawn on its validity and possible applications.

Chapter 2

Theories of emotions

2.1. Somatic basis

2.1.1. Brain structure

The brain can be divided into the neocortex, the surface of the brain, and the subcortical structures, including striatum and the brainstem. Several subcortical structures relate to event reward processing, especially those involving dopamine, and to reinforcement learning. They have connections with parts of the frontal and temporal lobes that are also involved in valuation. The cortical areas include the frontal cortex's medial and orbital regions (the inner surfaces and base of the frontal lobes), the amygdala (the inner surface of the temporal lobes) and the insular cortex (the junction of the frontal and temporal lobes) which, together with their subcortical counterparts, are referred to as the limbic system, shown in Figure 2.1. (Cohen 2005)

The limbic system is hard-wired with emotions and it is shared with other mammals. It regulates basic functions including body temperature and drives e.g. hunger, sleep and sex. Inside the limbic system, the thalamus receives sensory information and contains the emotional center amygdala, which is distributed on both sides of the brain. The limbic system normally passes information onto the cortex where higher cognitive functioning takes place, but it can intercept the signal by releasing catecholamine neurotransmitters which make it possible to react quickly to threats. (McDermott 2004, Cohen 2005)

This means that emotional processing partly takes place before cognition although it is important to realise that both systems are interconnected, the

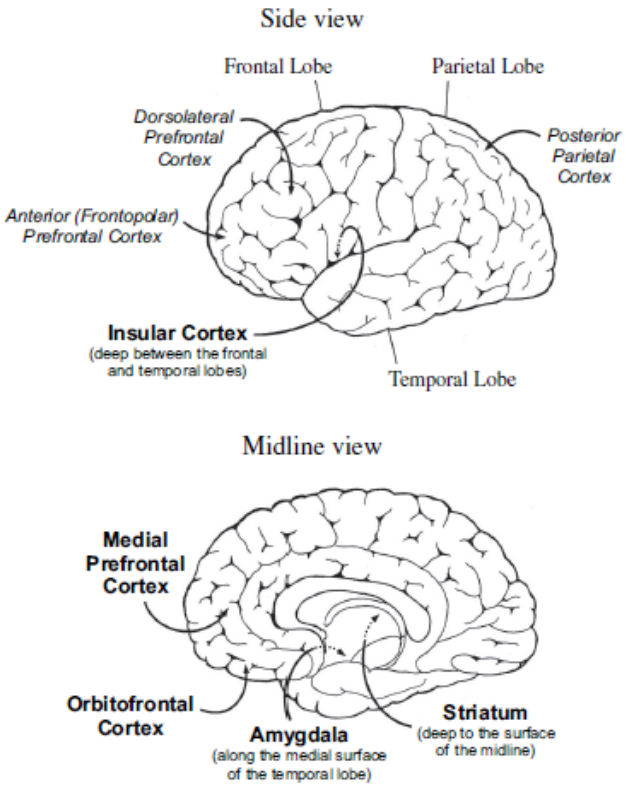


Figure 2.1: Brain areas related to decision making, bold text indicating areas related to emotional processing and italics to cognitive processing (Cohen 2005)

other direction being for example event appraisal. Especially the anterior and dorsolateral regions of prefrontal cortex, located in the upper and front most surfaces of the frontal lobes, are involved in higher-level cognitive processes. (McDermott 2004, Cohen 2005)

The concept of limbic system has been challenged because it has been hard to define in terms of an exact location as emotions seem to be associated with multiple brain areas. LeDoux (1995) reviewed neural activity related to fear conditioning, or Pavlovian defence conditioning, where an unconditioned stimuli (US) such as pain is preceded by a conditioned stimuli (CS) e.g. a sound. The stimuli tend to be associated only after a few pairings and the theory states that pathways transmitting information about the stimuli intersect. Research has shown that lesions of midbrain and auditory pathway's thalamic stations prevent conditioning, but those of auditory cortex do not, which implies that CS exits the auditory system at thalamus level. Indeed, in

addition to auditory cortex, thalamus projects to the amygdala and interrupting that connection also interferes with conditioning. The amygdala is crucial for CS conditioning. (LeDoux 1995)

Contextual conditioning, which pairs the US with the background stimuli present in the environment, is also dependent on the hippocampus. Logically, the hippocampus is involved in complex information processing. While lateral nucleus of amygdala is needed for input capture in conditioning, the cental nucleus is needed for its output. Lesions in the central nucleus reduce expression of conditioned responses. Other types of fear conditioning, e.g. visual, also involve amygdala but input circuit are not as clear as for auditory conditioning, see Figure 2.2. (LeDoux 1995)

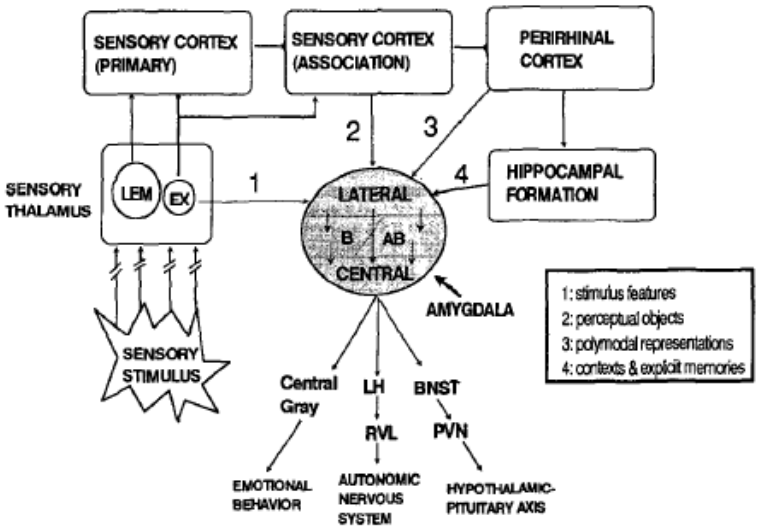


Figure 2.2: Neural circuits of fear conditioning (LeDoux 1995)

The neurological circuits provide clues for the cognitive-emotional debate in psychology (Zajonc 1984, Lazarus 1984). The amygdala receives input from both the thalamus and the cortex. The thalamus link is faster but not as detailed as the slower cortex link, which is needed to distuingish between stimuli. Inputs from hippocampal formation help to understand the context and raise different reaction to stimuli depending on the situation. The amygdala also sends some output to cortical sensory systems which allows emotions to influence perceptions. Feedback to the cortex, the hippocampus and the nucleus basalis can affect selective attention, spacial behaviour, contextual processing, memory and more. Emotions can be affected without complex information processing in the cortex and the hippocampus, which suggests that cognition is not a precondition for emotion. (LeDoux 1995)

Frank & Claus (2006) developed a model based on the interaction of the basal ganglia-dopamine (BG-DA) system and the orbitofrontal cortex (OFC) in decision making. BG-DA is involved in action selection and reinforcement learning, and OFC is critical for adaptive decision making and reversal learning. BG-DA integrates outcomes to create probability models for choices and OFC is used to provide working memory information on the magnitudes. The neural network model successfully simulates the effects of OFC damage on decision making. Wallis (2007) provides an excellent review of the functions of OFC.

Vytal & Hamann (2010) reviewed different studies of brain activation for discrete emotions. Activation likelihood estimation (ALE) was used because it allowed precise comparison of activation coordinates unlike brain region methods. Happiness, sadness, anger, fear and disgust were found to have characteristic activation patterns and their respective major regions were the right superior temporal gyrus, the left medial frontal gyrus, the left inferior frontal gyrus, the left amygdala, and the right insula and right inferior frontal gyrus. Appendix A shows a figure of the activation areas for the emotions.

2.1.2. Somatic marker hypothesis

The somatic marker hypothesis (SMH) states that decision making is influenced by marker signals created by bioregulatory, e.g. emotional, processes. Defect in emotion leads to impaired decision making. Emotional changes are collected in the term 'somatic state' because they are perceived as changes in the activity patterns of somatosensory structures. Somatic then refers to visceral and muscoskeletal aspects of the soma. SMH is based on an assumption of a complex human decision making process, where unconcious and concious layers interact. Cognitive operations are based on sensory information from the cortices and depend on supportive processes such as attention, working memory and emotion. Decision making uses dispositional information stored in the higher-order cortices and the subcortical nuclei. (Bechara et al. 2000)

The ventromedial prefrontal cortex (VM) stores dispositional linkages between knowledge and bioregulatory states. Links are learned to associate situations with bioregulatory states, including emotion, from past experiences. They are called dispositional because they are able to reactivate emotions in similar situations. The generated somatic states operate as incentive signals for option-outcome pairs in decision making situations. Figure 2.3 shows how normal people learned about the differences between the good and bad decks in a repeated gambling task, unlike the VM patients. Basically, the SMH

means that linking emotional content with alternatives is vital to smart decision making. (Bechara et al. 2000)

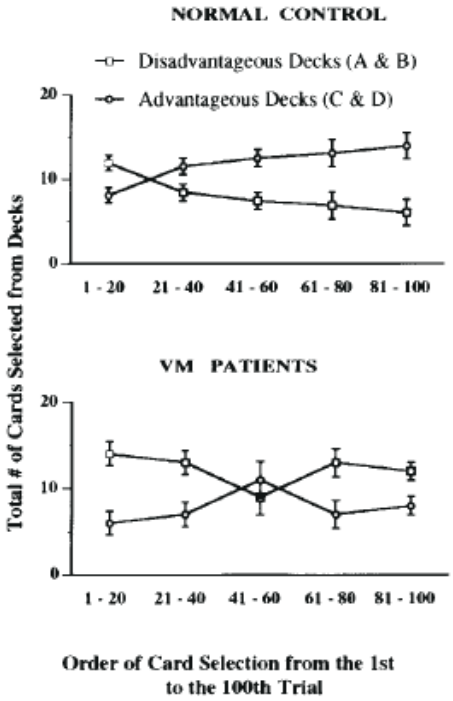


Figure 2.3: Differences between choices of gambling decks of normal people and VM patients (Bechara et al. 2000)

2.2. Psychological models

Scherer (2000) reviews psyhological models of emotion. He starts off by examining the definition of emotion and concludes that it is an episode of coordinated changes in several components. Some restrict the components to just subjective feelings but also physiological arousal and motor expression are commonly included. Emotions are often thought of as relevance detectors because they are triggered by significant events, either external or internal. Then other affective phenomena are compared to emotion. We will return to definitions of different types of affect in Section 2.2.2.

The major discussions in the psychology of emotions have a long history. Plato suggested a tripartite construction for the soul including separate areas of cognition, emotion and motivation. The debate about whether the

systems are separate and how they could be connected has been revitalized through the cognition-emotion debate (Zajonc 1984, Lazarus 1984). In Section 2.1.1 we noted that neurological evidence suggests an interconnected system. Descartes insisted that mental and physiological processes are examined at the same time. There are debates on whether emotions can be recognized with physiological patterns, and on how bodily changes affect subjective feeling, but physiological effects are not very relevant when dealing with robots. Darwin placed strong emphasis on universal expression of emotion in face, body and voice, which anthropologists have attacked arguing for the importance of cultural effects on emotion elicitation. There seems to be significant universality in emotions, see e.g. Ekman & Friesen (1971) and Levenson et al. (1992), so sociocultural factors are ignored in our model. (Scherer 2000)

2.2.1. Dimensions of emotions

A central debate in the psychology of emotions is whether there are basic emotions that can be mixed to represent other emotions or are there certain dimensions along which all emotions can be represented. Russell & Barrett (1999) dismiss the idea of basic emotions because the research does not converge on the correct set, as pointed out by Ortony & Turner (1990) (see Appendix B), and remark that languages have different amounts of emotion categories, e.g. 7 in Chewong to 500-2000 in English. Instead, they choose a circumplex model with two bipolar dimensions used for representing core affect, or mood. The dimensions are pleasure and arousal, and they are supported by factor analyses. Watson, Clark & Tellegen (1988) offer a similar structure with the dimensions positive and negative affect, which are basically a 45 degree rotation of pleasure and arousal, and describe the same space. However, for emotional episodes, more dimensions are needed. For example, anger and fear could fall in identical places in the circumplex but additional dimensions could distinguish them. Event appraisal (Roseman et al. 1990) can be used for this purpose. Emotional episodes involve core affect but are complex processes and often concerned with a specific object. The circumplex can be seen in Figure 2.4. (Russell & Barrett 1999)

Several researchers take an opposing stand and claim that emotions can be divided into basic categories which have separate qualities. For example facial expressions, language and brain activation methods have been popular in categorization (Ekman & Friesen 1971, Johnson-Laird & Oatley 1989, Izard 1994, Damasio, Grabowski, Bechara, Damasio, Ponto, Parvizi & Hichwa 2000, Phan, Wager, Taylor & Liberzon 2002). An important argument is that two-

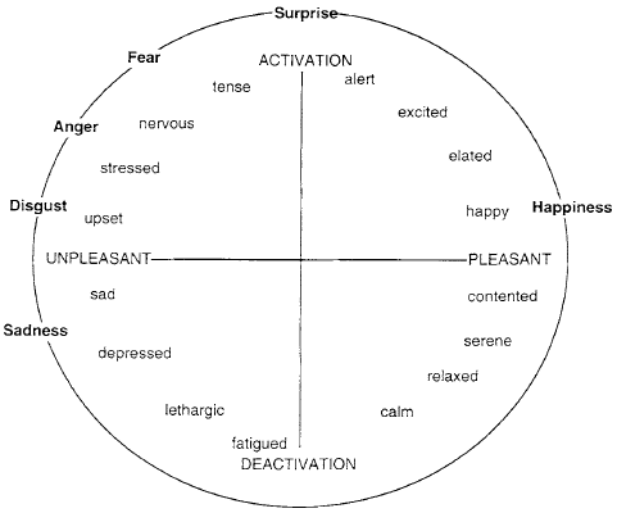


Figure 2.4: Russell's two-dimensional model with discrete emotions plotted (Russell & Barrett 1999)

dimensional bipolar models do not support mixed exclusive feelings, that is, one cannot e.g. be happy and sad at the same time. Larsen & McGraw (2001) showed that after watching the film *Life Is Beautiful*, happiness and sadness did co-occur strongly in the participants. Table 2.1 lists studies on discrete emotions. Happiness, sadness, anger, fear and disgust are identified in almost all the studies. It is useful to measure the emotional state as a mix of discrete emotions because specific emotions affect decision making in different ways (Lerner & Keltner 2000).

It is notable that the dimensions of emotion generation, identification and expression are not necessarily the same. An emotion could be generated through a motive-inconsistent and other-caused event, it could be identified as a mix of anger and disgust, and it could be expressed through aggressive facial, skeletal, vocal and autonomous activity. Emotion identification with a circumplex is inferior to using discrete categories, because although the latter can be mapped to the former, it does not work the other way so well.

But how to measure emotion generation? Izard (1993) lists neural, sensorimotor, motivational and cognitive systems as emotion activators. The neural system is too complex to model easily, and the sensorimotor system that e.g. associates postures with emotions is not so relevant for a robot. The motivational system includes physiological drives such as hunger and the cognitive system is related to perception of events. Low battery level can be chosen as a drive comparable with nutritional deprivation, or hunger, and similarly, it

Model developer	Emotions	Methodology
Ekman & Friesen (1971)	Happiness, Sadness, Anger, Fear, Surprise, Disgust	Facial expressions
Levenson et al. (1992)	Happiness, Sadness, Anger, Fear, Disgust	Autonomous nervous system activity
Frijda et al. (1989)	Joy, Sadness, Anger, Fear, Surprise Regret, Relief, Hope	Clustering of appraisal and action readiness components
Izard (1994)	Joy, Sadness, Anger, Fear, Surprise, Disgust Shame, Interest	Facial expressions
Ekman (1992)	Happiness, Sadness, Anger, Fear, Surprise, Disgust-Contempt	Meta-analysis
Johnson-Laird & Oatley (1989)	Happiness, Sadness, Anger, Fear, Disgust	Linguistic
Damasio et al. (2000)	Happiness, Sadness, Anger, Fear	Brain activity based grouping
Phan et al. (2002)	(Happiness), Sadness, (Anger), Fear, (Disgust)	Brain region activation meta-analysis
Vytal & Hamann (2010)	Happiness, Sadness, Anger, Fear, Disgust	Activation likelihood estimation meta-analysis

Table 2.1: Different models of discrete emotions

will increase anger.

On the cognitive side, Roseman et al. (1990) introduced a revised version of event appraisal. Event appraisal is based on the assumption that emotions are activated not by events as such, but rather by their perception on several dimensions. Empirical support was acquired for the effects of:

- situational state: whether the event is motive-consistent or motive-inconsistent
- motivational state: whether the subject is appetitive for a reward or aversive to a punishment
- agency: who is responsible for the event (self, other, circumstances)
- probability: how likely the event is
- power: perceived degree of control over the event
- legitimacy: deserving a positive or a negative outcome

The appraisal dimensions were tested in experiments where subjects were asked to recall experiences of two given emotions and tell the stories of what happened. Then, three questions regarding each appraisal dimension were asked in a random order and on a 9-point scale. The results in Appendix C show that all dimensions were significant, but probability was not as important as the other dimensions. Legitimacy was a controversial dimension and it does not appear in Figure 2.5 which visualizes the effects of appraisal dimensions on emotion activation.

Situational state differentiates positive emotions from negative ones. Motivational state differentiates joy and sadness from relief and disgust. The results imply that other-agency causes surprise, liking, disgust and anger/frustration, self-agency causes pride, shame, guilt and regret, circumstances cause hope, relief, sadness, and all cause joy, regret and fear. The causes of surprise, disgust, frustration, joy, relief, hope, distress, sadness and fear can be easily attributed to self or other people, even when the events are caused by circumstances, as when frustration in a bad grade for a difficult course is seen as caused by the teacher. Uncertain events elicit surprise, hope and fear, certain events elicit joy and disgust, and both kinds of events can elicit relief. Power is generally seen as higher in positive than negative emotions but this might be due to feeling powerful when positive results are achieved, and it is not necessarily a relevant dimension as it correlates with the situational state. (Roseman et al. 1990)

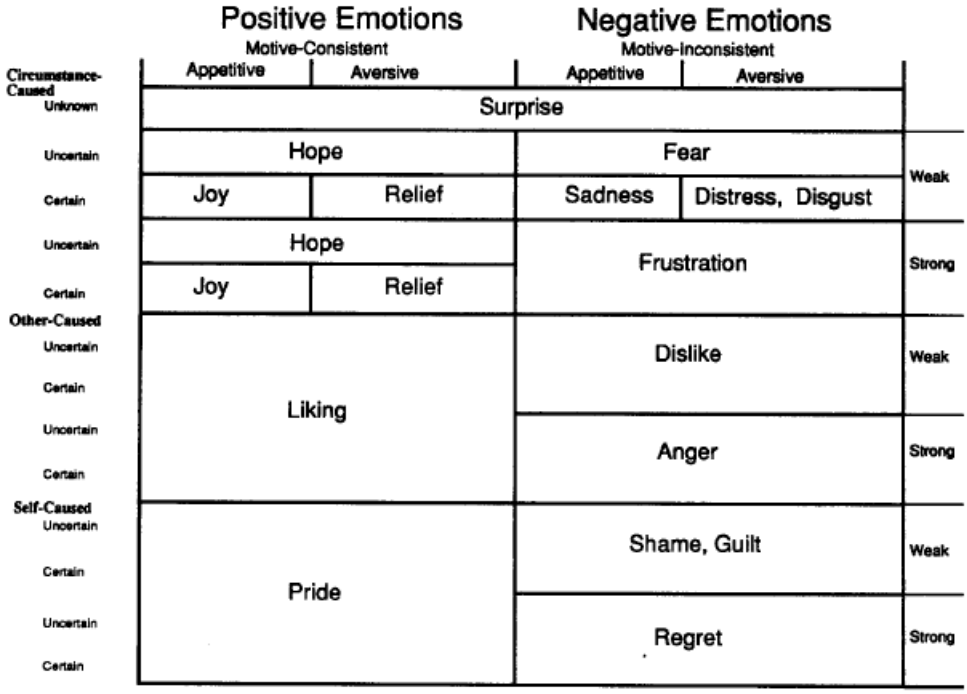


Figure 2.5: Roseman's event appraisal model (Roseman et al. 1990)

2.2.2. Core affect, emotional events and moods

A remarkable problem in the psychology of emotions is defining affective phenomena. Russell (2003) provides a framework for different levels of affect. *Core affect* is the most primitive and universal conscious level of affect, or the feeling component of other levels of affect, and it is measured on two dimensions; pleasure (or valence) and arousal (see Figure 2.4). It is a mental but not a cognitive concept as its cause is not necessarily defined, it simply is. Changes in core affect often guide attention towards their causes and like-valenced elements, which is seen as *mood congruency* in cognitive processing. Mood congruency refers to the focus of memory and other information processing on entities with similar valence as the current core affect. Core affect dynamics can be influenced by *affect regulation* actions e.g. drinking coffee to increase arousal. People normally try to maximize pleasure with their behavior and core affect is involved in motivation, reward and reinforcement. *Mood* is defined as a prolonged core affect.

Another important concept is *affective quality*. Objects are affectively interpreted before they are processed in consciousness. The affective quality of an object is a property which has the capacity to change core affect but does

not necessarily do so. Perception of this property is an evaluation process. Affective quality is closely related to core affect, but it requires an object. *Attributed affect* links the changes in core affect to a cause or an object. *Emotional episodes* are events that are sufficiently similar to certain emotional categories which are defined by multi-dimensional cognitive structures, e.g. becoming afraid or having fear when seeing a bear in the forest. Often an antecedent event contains affective qualities that change core affect, which is then attributed to the event. A complex appraisal follows, expressive changes appear, subjective conscious experiences, e.g. indecision, are felt, an emotional label is given to the episode and some action is taken. Emotional episodes can be mapped to the circumplex, but the intensities vary and e.g. Reisenzein (1994) suggested that emotions are not linear on the circumplex with regard to their intensities. (Russell 2003)

Beedie et al. (2005) conducted a questionnaire of the distinctions between emotion and mood as well as a review of the available academic studies. The answers to the open-ended question "*What do you believe is the difference between an emotion and a mood?*" were analyzed for content. Cause, duration and consequences are important criteria for both academics and non-academics, but intentionality and control are emphasized more by the academics and non-academics, respectively. The results are summarized in Figure 2.6. Mood is considered longer lasting than emotion and having less specific causes. Mood is seen as the result of emotions and it biases cognition while emotion biases behaviour. Non-academics mentioned having less control over emotions, and moods being experienced in thinking and emotions being experienced as feeling.

In this thesis, the term emotion will be used instead of emotional episodes and mood refers to the slowly changing, emotion-integrative core affect but not the core affect dimension of emotions.

2.2.3. Conditioning and attitudes

Sometimes the affective qualities observed in objects can raise attitudes if they appear constantly. Then, affective perception is biased by a conditioned emotion. Pavlov (1927) presented famous experiments on dogs where the animals showed conditioned defense reflexes for signals (conditioned stimuli, CS) which had been associated with harmful stimuli (unconditioned stimuli, US). He called it defense conditioning and today it is known as fear conditioning. Conditioned responses are rather emotions that affect behaviour than direct behaviour itself, for example in response to an object that is associated with

Criterion	*Non-academic (%)*	*Academic (%)*
Cause	65	31
Duration	40	62
Control	25	–
Experience	15	–
Consequences	14	31
Display	14	–
Intentionality	12	41
Anatomy	11	–
Intensity	11	17
Timing	8	–
Function	7	18
Physiology	7	8
Stability	7	–
Awareness of cause	4	13
Clarity	3	–
Valence	3	–

Figure 2.6: Comparison of distinctions between emotion and mood (Beedie et al. 2005)

joyful experiences, a subject would feel more happy instead of just acting in a specific happy manner. The neural basis for fear conditioning was reviewed in Section 2.1.1. (LeDoux 1995)

Attitudes can be understood as evaluative tendencies for affective qualities. Several studies have shown that conditioning can have effects on attitudes, e.g. Olson & Fazio (2001) proved that pairing positive and negative words with Pokemon names affected the valence ratings of the characters. Also subliminal stimulus have been observed to affect evaluation which suggests that unconcious conditioning is effective for attitudes. The results of Öhman & Soares (1998) suggest that we are more prepared to make unconcious conditions for US-CS pairs with evolutionary or cognitive basis, e.g. a shock can be conditioned unconciously with spiders but not with flowers. Often attitude also changes through evaluative association which differs from classical conditioning in that the US and CS are presented at the same time. This kind of association endures even if the CS appears without the US, which usually leads to extinction in classical conditioning. Attitudes can be drawn from associated stereotypes when individuals are seen as part of a group that has been attributed with affective qualities. As well as affecting the evaluation of events, attitudes can elicit emotions. (Clore & Schnall 2005)

2.2.4. Expectations and alternatives

Situational state, or motive-consistency, of events is often measured in computational models through utility. Decision Affect Theory (DAT) (Mellers et al. 1997), FLAME (El-Nasr et al. 2000) and EMA (Gratch & Marsella 2004) are examples of this approach. The higher the utility of an event, measured with several criteria, the higher the motive-consistency. However, perception of motive-consistency is also affected by expectations of what should happen and alternative scenarios. Roboceptionist (Kirby et al. 2010) changes its mood according to not only the present situation, but to memories of past events and expectations of future events too. The uncertainty dimension of Roseman et al. (1990) is specifically related to future events, which cause surprise, hope and fear. The current visceral state, which refers to the mix of emotional and drive states, affects perceptions more than remembered or expected states because people are bad at imagining how they would value options in other states (Loewenstein 2000).

Other alternatives are powerful in changing perceptions of a decision problem. Winning zero in a gamble can elicit both positive and negative emotions if the other outcomes would have been negative or positive (Mellers et al. 1997). This suggests that utility is perceived as relative rather than absolute.

2.2.5. Personality traits for diverse dynamics

People are very different in terms of behaviour, feeling and thinking but usually some enduring and similar patterns can be recognized which are called a personality. Properties of a personality, or personality traits, can be described with factor models. The most common one has five factors and it is called The Big Five or OCEAN which stands for Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. The factors correlate with groups of properties in Table 2.2. The Big Five is consistent across cultures and works as a framework for also abnormal personalities. Other factor models with two to five factors have been proposed and examining them reveals the hierarchical structure of the factors. Figure 2.7 shows correlations between the different factors. (McCrae & P. T. Costa 1997, Markon et al. 2005)

Personalities are important for emotional models because they affect the dynamics of emotion activation together with mood. There are differences in average level, variability, activation and intensity of emotions. For example the TAME model uses personalities in activation functions (Moshkina 2006).

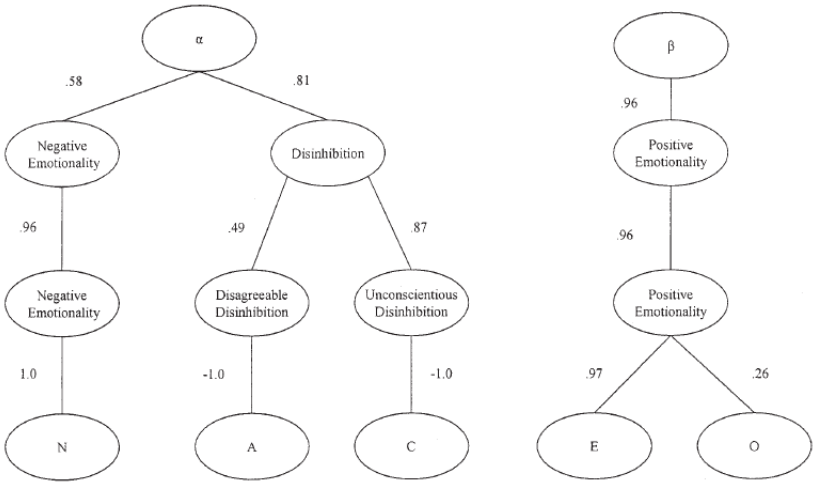


Figure 2.7: Markon's structural model of personalities, N=Neuroticism, A=Agreeableness, C=Concentiousness, E=Extraversion, O=Openness (Markon et al. 2005)

Studies have found different and sometimes contradictory results for the effects when personality and mood have been examined separately, but instead of this traditional approach, moderation and mediation frameworks seem functional. In moderation, personalities and moods interact in influencing emotional processing, and in mediation, personality traits guide toward specific emotions and moods that influence emotional processing. Most importantly, neuroticism is connected to higher intensities and variation as well as tendency towards negative-valenced emotions whereas extraversion is related to positive-valenced emotions. (Rusting 1998, Eid 1999)

Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Fantasy, Aesthetics, Feelings, Actions, Ideas, Values	Competence, Order, Dutifulness, Deliberation, Self-Discipline, Achievement Striving	Warmth, Gregariousness, Assertiveness, Activity, Excitement Seeking, Positive Emotions	Trust, Modesty, Altruism, Compliance, Straightforwardness, Tender-Mindedness	Anxiety, Impulsiveness, Depression, Vulnerability, Angry Hostility, Self-Consciousness

Table 2.2: Correlates of personality factors (McCrae & P. T. Costa 1997)

Following the mediation approach, personality traits make people more sensitive to certain emotional events. Eid (1999) conducted an experiment on how personalities affect the variability of discrete emotions. The results in Figure 2.8 show that neuroticism was significantly correlated with mean levels and standard deviations of all negative (positive correlation) and almost all positive (negative correlation) emotions. Extraversion significantly affected mean levels of all but love and had positive correlation for both mean levels and standard deviations of positive emotions and vice versa for negative emotions. Agreeableness was associated with higher mean love and happiness as well as lower anger and sadness. Openness did not affect anything and conscientiousness only correlated with higher levels of happiness.

Basic category	Neuroticism	Extraversion	Openness to Experiences	Agreeableness	Conscientiousness
Love (N = 169)					
M	−.17*	.12	.04	.27**	.09
SD	.20*	.00	.01	−.04	−.10
Happiness (N = 170)					
M	−.42**	.26**	−.02	.29**	.17*
SD	.14	.02	.00	.02	.00
Joy (N = 170)					
M	−.25**	.35**	−.04	−.02	.12
SD	−.08	.25**	−.05	.04	.03
Anger (N = 170)					
M	.35**	−.19*	−.05	−.19*	−.09
SD	.32**	−.05	−.05	−.18*	−.05
Fear (N = 169)					
M	.47**	−.20*	.00	−.08	.01
SD	.45**	−.009	.06	−.03	−.03
Shame (N = 170)					
M	.31**	−.17*	.00	−.11	−.15
SD	.38**	−.05	.01	−.08	−.14
Sadness (N = 170)					
M	.46**	−.22**	.00	−.17*	−.12
SD	.49**	−.21**	.02	−.08	−.13

* $p < .05$. ** $p < .01$.

Figure 2.8: Correlations between personality and mean level and standard deviation of emotions (Eid 1999)

Heller (1993) proposed a model linking personality with affect dimensions based on brain activity. Neuroticism is associated with low pleasantness as high activation and extraversion is associated with high pleasantness and high activation. Low neuroticism is often called emotional stability and low extraversion is called introversion, and they have opposite associations as shown in Figure 2.9. Schmidtke & Heller (2004) tested emotional brain activation of different personalities with electroencephalography (EEG) and concluded that the results, together with cited research, supported the model. It can be argued that a neurotic person would be more biased towards anger than sadness and an extroverted person more towards joy than happiness because activation is higher in anger and joy than in sadness and happiness.

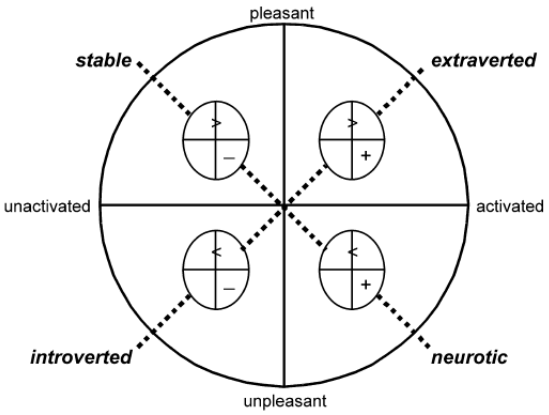


Figure 2.9: The circumplex structure of personality by Heller (1993) (Schmidtke & Heller 2004)

Kuppens et al. (2010) developed a dynamic stochastic differential equation model, DynAffect, for affect dynamics in the circumplex. They based their model on individual differences in affective variability and a homebase with an attractor strenght. The empirical data was fitted with the model which provided similar affect patterns as the subjects. Several variables were examined regarding their effect on the parameters: self-esteem, neuroticism, extraversion, normal positive and negative affect, satisfaction with life, reappraisal, suppression and rumination. The correlations can be seen in Figure 2.10. The average homebase had slightly positive pleasure and arousal, and high neuroticism moved it towards higher arousal and lower pleasure, whereas extraversion only moved it significantly towards higher pleasure.

Dispositional measure	Valence home base	Arousal home base	Valence variability	Arousal variability	Valence attractor strength	Arousal attractor strength
Neuroticism	**−.386** (.000)	**.230** (.042)	**.278** (.013)	−.046 (.688)	−.098 (.392)	−.175 (.122)
Extraversion	**.288** (.010)	−.197 (.083)	−.003 (.979)	.156 (.171)	.034 (.766)	.085 (.457)
Positive affect	**.446** (.000)	−.125 (.272)	.025 (.830)	.097 (.394)	.064 (.573)	.117 (.304)
Negative affect	**−.227** (.044)	.104 (.360)	**.283** (.011)	−.098 (.388)	.076 (.503)	.065 (.567)
Self-esteem	**.356** (.001)	**−.262** (.019)	**−.253** (.024)	.025 (.826)	.058 (.614)	.144 (.207)
Self-esteem variability	.069 (.548)	−.105 (.358)	**.393** (.000)	**.242** (.032)	.158 (.165)	.096 (.398)
Satisfaction with life	**.247** (.028)	−.178 (.116)	−.185 (.103)	.061 (.595)	.066 (.562)	.058 (.614)
Reappraisal	.034 (.763)	.014 (.901)	−.212 (.061)	−.120 (.292)	.160 (.159)	**.310** (.005)
Suppression	−.146 (.199)	**.230** (.008)	.125 (.273)	−.035 (.760)	−.017 (.879)	−.034 (.764)
Rumination	−.080 (.482)	−.042 (.711)	.068 (.552)	−.177 (.118)	−.011 (.925)	**−.226** (.045)

Figure 2.10: DynAffect model correlations between dispositions and parameters with p-values in parentheses (Kuppens et al. 2010)

2.2.6. Human needs and subjective well-being

What motivates humans in their behavior? Maslow (1943) introduced a hierarchy for human needs that guides behavior. It is organized in ranked goal classes and prior needs must be satisfied before concentrating on a new goal although many needs can motivate behavior simultaneously. Humans seek to satisfy the needs one by one and move up in the hierarchy. The classes are in the following order:

1. Physiological needs
    - food and water
    - air for breathing
    - sex and reproduction
2. Safety needs
    - personal security
    - financial security
    - health
3. Love and belonging
    - friendship
    - intimacy
    - family
4. Esteem
    - accepting oneself
    - receiving respect
5. Self-actualization
    - enhancing oneself
    - fulfilling desires

The model was developed for individual motivation and it has received both support (Wicker, Brown, Wiehe, Hagen & Reed 1993) and criticism (Hofstede 1984). Hagerty (1999) applied the model to nations' Quality of Life (QOL) by using approximate measures the human needs, e.g. daily calories and Gross

Domestic Product (GDP) for physiological needs. Results mostly confirmed the theory but external conditions such as environmental health or poverty in society are not considered well enough by the theory.

Subjective Well-Being (SWB) is another concept for understanding QOL, relying on subjective perception. Often the research concentrates on correlates of happiness which is mostly measured with life satisfaction, positive and negative emotions. Diener, Suh, Lucas & Smith (1999) reviewed the research on SWB and concluded that input variables affect the components differentially. Interestingly, on a national scale, income did not affect SWB significantly. Tay & Diener (2011) conducted a large-scale international review on SWB in various cultures. They tested the effects of several needs on life evaluation, positive and negative feelings. The needs were close to the hierarchy of Maslow (1943): Basic needs for food and shelter, Safety and security, Social support and love, Feeling respected and pride in activities, Mastery, Self-direction and autonomy, and Log income. Basic needs were the most influential predictor for life evaluation and they were partly explained by income, since income did not affect positive or negative feelings as it did life evaluation. The results supported Maslow in that basic and safety needs are often satisfied before other needs but they also showed that well-being can be enhanced with psychological needs even if basic needs are not met.

2.3. Emotions in economics

Economics has seen waves of trends and theories. It started out with classical economics and Adam Smith's model for self-interest and division of labor. Early theories were motivated by the effect of environment on behavior. Later Keynes' theory became widely used as it provided tools for economic management based on psychological concepts of behavior. Then came neoclassical economics with Samuelson's revealed preference model, which was based on strongly normative axioms for choice. Morgenstern and von Neumann created the concept of expected utility (EU) maximization and game theory for multiple actors. EU provided a useful framework for decision making. Soon, however, evidence became to amount against axioms and EU models in real behavior. Friedman supported neoclassical models claiming that even if the underlying behavioral theories were wrong, the models' predictions could still be true in an approximate sense. Morgenstern concluded that EU was applicable in a limited domain. (Glimcher, Camerer, Fehr & Poldrack 2008, Ch. 1)

However, research evolved into a fresh direction. The Allais paradox in which serial pairwise choices led to a revealed preference which violated the independence axiom of EU, and psychological experiments where the "framing" effect showed that choice is dependent on the description of alternatives, gave birth to behavioral economics. The idea was to use psychological evidence to improve neoclassical theories. Kahneman and Tversky were pioneers creating prospect theory for the effect of references on choice. Observations of choice were anticipated to produce heuristics for statistical models and social preference theories to add the impact of other people's values on choice. Experimental economists such as Plott and Smith were more interested in finding global rules for economics using methods from psychology for experiments as opposed to using psychological principles to enhance models. (Glimcher et al. 2008, Ch. 1)

Neuroeconomics was founded because of the need to get new evidence on information processing mechanisms in humans. Brain research could bring new knowledge to enhance algorithms developed by behavioral economists. Neuroeconomists were split between two approaches; the behavioral economic, using economic theory to develop choice algorithms, and the neuroscientific, trying to improve neoclassical models with new research with neurological tools. Neuroscientific behavior is studied with physiological models which aim at correlating physiological signals, or changes in biological states, with behavior but the experiments are slow and often destructive. Another way is to use neurological methods. Neurological research concentrated on the effects of brain damage on behavior. Brain damage was known to affect also mental states, but they were harder to observe and were largely ignored. Signal detection theory provided a way to relate brain activity with behavior. Research with stochastic monkey choices suggested the use of psychometric-neurometric match, a correlation between behavior and neural activity, although it did not seem to suit all brain areas. An expected utility theory based on brain stimulation of rats and Herrnstein's matching law was proposed, normalizing utility with regard to other alternatives. (Glimcher et al. 2008, Ch. 1)

Later non-invasive techniques for brain imaging, funtional magnetic resonance imaging (fMRI), and for brain stimulation, transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS), have given interesting results because now also mental functions could be studied. fMRI was widely used to examine brain activity in cognitive tasks. An increasing amount of neuropeptide oxytocin was observed to alter behavior differently in human and nonhuman interaction for a trust game which implied that people have different mechanisms for social choice. The disruption of only the right

side of the dorsolateral prefrontal cortex (DLPFC) was shown to increase the acceptance of unfair offers in the ultimatum game even though both sides of the DLPFC are activated during the game. Brain imaging and stimulation can lead to understanding causal mechanisms. (Glimcher et al. 2008, Ch. 1)

There has been a growing interest in the role of emotion in economics as new research has highlighted its role in economic decisions. An important notion is that subjective experience, or feeling, is not necessary for emotions to be present. That is, we do not always know that our emotions are affecting our decisions. A problem in introducing the effect of emotions on decision making in economics is how definitions are used. Consumer preferences and value are compared to emotional responses to events and attitudes toward affective objects. Preferences are understood as long-lasting liking or disliking that are connected to emotional states in consumer research. Affect dispositions are described as enduring traits that could be compared to personality traits in psychology. In decision theory, preferences refer to personal utility function weights and shapes, and risk attitudes. (Glimcher et al. 2008, Ch. 16), (Clemen 1996)

2.4. Connection with game theory

Emotions are closely connected with game theory as was mentioned in the previous section. Expected utility (EU) maximization hypothesis and many axioms have been proved to be descriptively invalid. Probably the clearest example is the ultimatum game. In stage one, the first participant makes an offer of splitting a given amount of money between her and the other participant; in stage two the other participant either accepts or rejects the offer. If the offer is rejected, neither participant gets anything. According to utility maximization, using monetary gain as the single criterion, the first player should offer the smallest possible amount and the second should accept it, however, unfair offers which are less than 20% are mostly rejected. Of course, adding other criteria might overcome the problem. Sanfey et al. (2003) found that unfair offers increased brain activity in areas related to both cognition and negative emotions, specifically anger and disgust. (Camerer & Thaler 1995)

The dictator game is a simplified version of the ultimatum game where the first player (the dictator) has all the power and second player (the recipient) does not have a choice to reject. Even in this game the average give ratio is 28%. Different experiment environments affect the distribution but zero of-

P1 \ P2	C	D
C	2,2	0,3
D	3,0	1,1

Table 2.3: The utilities in prisoner's dilemma, utility pairs a,b representing the utilities of player 1 (P1) and player 2 (P2), respectively

fers are very rare. The individual give distribution has spikes in 0-9%, 45-54% and 91-100% segments. Old age increases, and the perception of deserving a reward decreases, the give rate the most. Probably giving elicits positive emotions which are regarded as beneficial. (Engel 2010)

In the prisoner's dilemma, two players decide whether they will co-operate (C) or defect (D) without knowing each others' actions. Each player has an incentive to defect regardless of the other player's choice but if both players defect, they will get less than otherwise, as shown in Table 2.3. This situation gives an equilibrium of both defecting. In a repeated game some strategies can lead to a co-operation equilibrium. In practice the choice is not always clear, as when one player knows that the other has defected but is feeling empathy towards her and decides to co-operate even though it is against self-interest. (Batson & Ahmad 2001)

The trust game is in a way the opposite of the ultimatum game. A truster can send part of a given amount of money to a trustee and that money is then multiplied and the trustee can return part of the money. According to game theory, the truster shouldn't give anything. Berg, Dickhaut & McCabe (1995) found that the trusters sent over 50% on average in a simple design, and based on various experiments they concluded that reciprocity is a basic human element. Engle-Warnick & Slonim (2004) showed that in repeated games experienced players did give less when the relationships were definite compared to indefinite relationships. Delgado, Frank & Phelps (2005) studied brain activation in trusters with given perceptions of trustees (good, bad, neutral) and noticed that the trusters made riskier choices with good trustees. Also, the repeated games and feedback were much more efficient for learning about a neutral than a good or bad trustee.

For multiple players, public good games are another example where categorization of humans into types clarifies the setting. In them, the players can choose how much they want to invest in the public goods that benefit everyone and how much on themselves. The players can be divided in three types: *i*) cooperators, who focus on investing in public goods, *ii*) free-riders, who invest in themselves, and *iii*) reciprocators, who use a strategy conditional on beliefs

about the distribution of types. The population as a whole is a stable polymorhic equilibrium of types, probably due to evolutionary and social causes. The public goods contributions are then affected by compositions of the groups in which the games are played. (Kurzban & Houser 2005, Gunnthorsdottir, Houser & McCabe 2007)

One suggestion for making EU work in these contexts is the introduction of types and beliefs about them. Every type of individual is assumed to use a certain strategy and beliefs on the distribution of types in the group affect the optimal choice. These beliefs are affected by emotions. (Glimcher et al. 2008, Ch. 5)

2.5. Effect of emotions on decisions

Affect influences judgement diversely. For example, happy audiences feel more positive about presented messages, fearful people perceive also other people as more fearful and depression causes decreased attraction to others. The Affect Infusion Model (AIM) seeks to explain mood influences on social judgements. Often the effects are related to mood congruency e.g. selective attention. Affect has significant weight when heuristic processing is used in the judgements. Unfamiliarity, simplicity, personal relevance, low motivation, low cognitive capacity and positive affect lead to heuristic processing and high affect infusion which refers to the effect of emotion on judgement. (Forgas 1995)

Schwarz (2000) reviewed effects of emotions on decisions and noted that happy people engage in heuristic and sad people in systematic processing strategies, confirmed by de Vries, Holland & Witteman (2008). The heuristic and systematic processing strategies can be compared to the System 1 and System 2 thinking modes of Kahneman (2011). Happy people are also more imitative in multiplayer games whereas sadness evokes analysis. Pham (2007) notes that angry and disgusted people also use heuristic processing which suggests that the valence dimension is not the only predictor of evaluation strategy. Tiedens & Linton (2001) claimed that the certainty dimension of appraisal is important in determining the processing strategy, e.g. uncertainty-induced fear promotes analytic processing, unlike disgust. Anticipated regret and disappointment influence decisions when future emotions are predicted but unfortunately predictions are often biased by the current affect status. Affect influences are larger when the affect attributed to a specific object rather than to chance. Another observation was that for past experience, the peak

intensities and ending affects were much more significant than the duration. (Schwarz 2000)

Decisions can be based on maximizing not the expected utility (EU), but the anticipated positive emotions or valence instead. Both immediate and expected emotions then influence decisions. Future utilities are discounted hyperbolically in the optimization of behavior. The immediate emotions, together with the other visceral factors, influence perceptions of risks and utilities as well as the criteria preferences. Loewenstein (1996) implies that the criteria weights are dependent on the visceral state. Drive states can dominate emotions in action selections. For example, very hungry people tend to focus on getting food regardless of being angry or happy. Tversky & Kahneman (1992) developed Prospect Theory to explain the anomalities of EU. They use nonlinear weighting functions for probabilities, as seen in Figure 2.11, and value functions which are convex below the reference point (negatively perceived outcomes) and concave above it, where the value functions are steeper for losses than for gains. (Loewenstein 2000, Loewenstein & Lerner 2003)

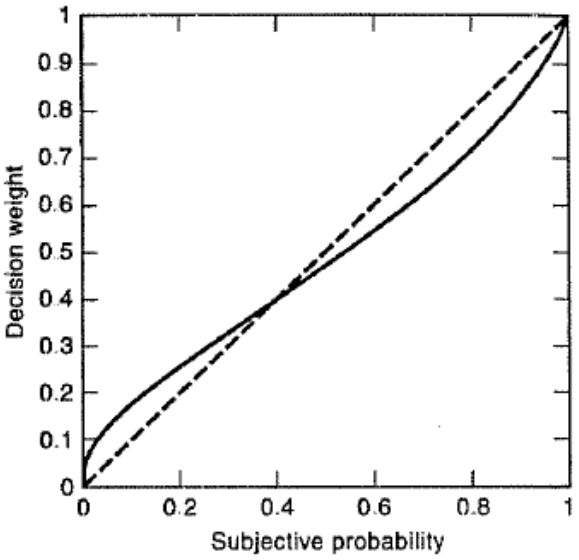


Figure 2.11: Probability weighting function inspired by Tversky & Kahneman (1992) (Loewenstein & Lerner 2003)

Lerner & Keltner (2000) argued that the effects of discrete emotions on decisions should be inspected rather than only considering the valence of the emotions. They chose to use an appraisal-tendency approach where certain event appraisal dimensions are used as the drivers of emotion perception effects, depending on the nature of the event. For example, anger and fear

that are both negative emotions, can affect risk perception in opposite ways. Especially in events that are ambiguous with regard to certainty and controllability, the effects of anger are similar to those of happiness while otherwise they are closer to those of fear (Lerner & Keltner 2001). Anderson & Galinsky (2006) support the claim that power has an effect on risk-taking. Raghunathan & Pham (1999) concluded that there are different effects for also anxiety (close to fear) and sadness. They proposed that sadness biases toward high-risk/high-reward options and anxiety toward low-risk/low-reward options because the goals are reward acquisition for sadness and control accretion for anxiety. Fessler, Pillsworth & Flamson (2004) took an evolutionary approach and found sex differences in risk taking for anger and disgust. They hypothesized that men are more risk-seeking in an angry state and women more so in a disgusted state because the emotions are associated with protecting reproductive interests. Risk behavior did not change when the emotions were in reverse order for the sexes and the effects decrease with age. The results of the studies are summarized in Table 2.4. The effect magnitudes are mostly related to emotion intensities.

Emotion	Risk perception	Processing strategy
Happiness	Optimistic	Heuristic
Anger	Optimistic when an ambiguous event or male, pessimistic otherwise	Heuristic
Disgust	Optimistic if female	Heuristic
Sadness	Optimistic	Analytic
Fear	Pessimistic	Analytic

Table 2.4: The effects of discrete emotions on risk perception

The observation that different emotions increase the use of heuristic or analytic processing strategies could suggest that various frameworks are needed for behavior selection. Decision theory is useful as an analytic environment but for heuristic processing, we would need something similar to somatic markers (Bechara et al. 2000) that intuitively point at the best action. El-Nasr et al. (2000) and Cattinelli, Goldwurm & Borghese (2008) used Q-learning in their model. It associates each (state,action) -pair with a number that reflects the goodness of that combination. The numbers are updated according to the outcomes.

Chapter 3

Modeling affective decision making

3.1. Decision model

The decision model uses multi-attribute utilities and adaptive probability models to choose robot actions according to expected utility. Model averaging is used to combine results and the different users (human interaction partners) are taken into account. The model is based on Esteban (2012).

3.1.1. Framework and sets

Agent A is controlled by the model and multiple human users $B_1, \ldots, B_r \in \mathcal{B}_u$ interact with A. The interaction takes place within an environment E. A has a finite action set $\mathcal{A} = \{a_1, \ldots, a_m\}$ and the users B_u have another action set $\mathcal{B} = \{b_1, \ldots, b_n\}$, shown in Table 3.1. The environment E, which is common to the agent and the users, has dynamic states within a set $\mathcal{E}$. The environment states are both internal and external, e.g. simulated pain and temperature.

$\mathcal{A}$	cry, alert, warn, ask for help, salute, play, speak, ask for playing, ask for charging, ask for shutting down, tell jokes, tell stories, tell events, obey, do nothing
$\mathcal{B}$	recharge, stroke, flatter, attack, offend, move, update, speak, play, order, ignore, do nothing

Table 3.1: The action sets for the robot and the users

A has q sensors to read both internal and external states. Sensor readings are made at times t and the sensor reading vector is $\boldsymbol{s}_t = (s_t^1, \ldots, s_t^q)$. The agent interprets the environment state vector $\mathbf{e}_t$ with a possibly probabilistic function f:

$$\hat{\boldsymbol{e}}_t = f(\boldsymbol{s}_t).$$

A assesses the probabilities of the users interacting with it with function h

$$\hat{B}_u^t = h(\boldsymbol{s}_t).$$

A also evaluates the actions of users with a possibly probabilistic function g

$$\hat{b}_t = g(\boldsymbol{s}_t).$$

In short, A uses sensors $\boldsymbol{s}_t$ to recognize users $\mathcal{B}_u$, interpret the state $\mathbf{e}_t$ and assess the users' actions b_t. Then the forecasting model is updated with the new information. The forecasting model is used in calculating expected utilities for different actions.

3.1.2. Forecasting models

The agent has a forecasting model for evaluating probabilities for future user actions and environment development using past actions of users and the agent (also current action for agent), and the evolution of the environment ($\mathbf{e}_{t-1}, a_{t-1}, b_{t-1}$).

The agent has a limited memory, in this case two steps. For each user B_u^t, the probabilities of current scenarios are calculated as

$$\begin{aligned} p(\mathbf{e}_t, b_t \mid a_t, (\mathbf{e}_{t-1}, a_{t-1}, b_{t-1}), (\mathbf{e}_{t-2}, a_{t-2}, b_{t-2}), B_u^t) = \\ p(\mathbf{e}_t | b_t, a_t, (\mathbf{e}_{t-1}, a_{t-1}, b_{t-1}), (\mathbf{e}_{t-2}, a_{t-2}, b_{t-2}), B_u^t) \\ \times p(b_t | a_t, (\mathbf{e}_{t-1}, a_{t-1}, b_{t-1}), (\mathbf{e}_{t-2}, a_{t-2}, b_{t-2}), B_u^t). \end{aligned} \tag{3.1}$$

The first term is called the *environment forecasting model*. The external environment is partially controlled by the users and not by the agent. For example, the users control the lighting and the temperature of a room but not outside. In any case they can plug in the bot for charging. Only the latest user actions can affect the evolution of the environment. We use

$$p(\mathbf{e}_t \mid b_t, a_t, (\mathbf{e}_{t-1}, a_{t-1}, b_{t-1}), (\mathbf{e}_{t-2}, a_{t-2}, b_{t-2}), B_u^t) = p(\mathbf{e}_t \mid b_t, \mathbf{e}_{t-1}, \mathbf{e}_{t-2}). \quad (3.2)$$

The second term is called the *user forecasting model* and is used for evaluating probabilities of the current user actions. The probabilities are assessed using historical actions and the current agent action as follows

$$p(b_t \mid a_t, (\mathbf{e}_{t-1}, a_{t-1}, b_{t-1}), (\mathbf{e}_{t-2}, a_{t-2}, b_{t-2}), B_u^t) = p(b_t \mid a_t, b_{t-1}, b_{t-2}, B_u^t). \quad (3.3)$$

The equation is divided into two independent models which give possibly differing evaluations. The first model uses historical user actions to utilize action patterns

$$p(b_t \mid b_{t-1}, b_{t-2}, B_u^t). \quad (3.4)$$

The second model uses the current agent action to explain the user action as a response

$$p(b_t \mid a_t, B_u^t). \quad (3.5)$$

Combining the models using model averaging (Hoeting, Madigan, Raftery & Volinsky 1999) gives

$$\begin{aligned} &p(b_t \mid a_t, b_{t-1}, b_{t-2}, B_u^t) = \\ &\Big[p(M_2 \mid B_u^t)\; p(b_t \mid b_{t-1}, b_{t-2}, B_u^t) + p(M_1 \mid B_u^t)\; p(b_t \mid a_t, B_u^t) \Big], \end{aligned} \quad (3.6)$$

where $p(M_i \mid B_u^t)$ is the probability of forecasting model M_i usage for B_u^t and $p(M_1 \mid B_u^t) + p(M_2 \mid B_u^t) = 1,\ p(M_i \mid B_u^t) \geq 0$. The probabilities describe the action mentality of the user and more models could be used to capture different effects.

The complete forecasting model becomes

$$p(\mathbf{e}_t, b_t \mid a_t, (\mathbf{e}_{t-1}, a_{t-1}, b_{t-1}), (\mathbf{e}_{t-2}, a_{t-2}, b_{t-2})) =$$
$$\sum_u \left[p(\mathbf{e}_t \mid b_t, \mathbf{e}_{t-1}, \mathbf{e}_{t-2}) \times p(b_t \mid a_t, b_{t-1}, b_{t-2}, B_u^t) \times p(B_u^t) \right]. \tag{3.7}$$

Conditional probabilities are used to forecast m steps ahead. For example, two steps is

$$\begin{aligned}
&p((\mathbf{e}_{t+1}, b_{t+1}), (\mathbf{e}_t, b_t) \mid a_t, (\mathbf{e}_{t-1}, a_{t-1}, b_{t-1}), (\mathbf{e}_{t-2}, a_{t-2}, b_{t-2})) = \\
&p((\mathbf{e}_{t+1}, b_{t+1}) \mid a_t, (\mathbf{e}_{t-1}, a_{t-1}, b_{t-1}), (\mathbf{e}_{t-2}, a_{t-2}, b_{t-2}), (\mathbf{e}_t, b_t)) \times \\
&p(\mathbf{e}_t, b_t \mid a_t, (\mathbf{e}_{t-1}, a_{t-1}, b_{t-1}), (\mathbf{e}_{t-2}, a_{t-2}, b_{t-2})) = \\
&\sum_u \left[p(\mathbf{e}_{t+1} \mid b_{t+1}, b_t, \mathbf{e}_t, \mathbf{e}_{t-1}, \mathbf{e}_{t-2}) \times p(b_{t+1} \mid a_t, b_t, b_{t-1}, b_{t-2}, B_u^{t+1}) \times p(B_u^{t+1}) \right] \times \\
&\sum_u \left[p(\mathbf{e}_t \mid b_t, \mathbf{e}_{t-1}, \mathbf{e}_{t-2}) \times p(b_t \mid a_t, b_{t-1}, b_{t-2}, B_u^t) \times p(B_u^t) \right].
\end{aligned} \tag{3.8}$$

3.1.3. Action selection

The agent maximizes the expected utility r steps ahead by choosing its actions. Expected utility for an action strategy is

$$\begin{aligned}
&\max_{(a_t, \ldots, a_{t+r})} \psi(a_t, \ldots, a_{t+r}) = \sum_{(b_t, \mathbf{e}_t), \ldots, (b_{t+r}, \mathbf{e}_{t+r})} \left[\sum_{i=0}^{r} u(a_{t+i}, b_{t+i}, \mathbf{e}_{t+i}) \right] \times \\
&p((b_t, \mathbf{e}_t), \ldots, (b_{t+r}, \mathbf{e}_{t+r}) \mid (a_t, a_{t+1}, \ldots, a_{t+r}, (a_{t-1}, b_{t-1}, \mathbf{e}_{t-1}), (a_{t-2}, b_{t-2}, \mathbf{e}_{t-2}))).
\end{aligned} \tag{3.9}$$

This can be solved through dynamic programming with Bellman's equation (Bellman 1957). The utility function can be modified to direct the agent toward a desired state using $u(c) - \rho(c, c^*)$ where ρ is a distance to the ideal consequence value c^*. To make the agent less predictable, action probabilities proportional to the power function of expected utilities can be used

$$P(a_t) \propto \psi(a_t)^{cf}, \tag{3.10}$$

where $P(a_t)$ is the probability of choosing a_t and cf is the consistency factor of the agent.

3.1.4. Objectives and utilities

The robot has five caterogies of objectives: being charged, being secure, being taken into account, being accepted and being updated. These are comparable with the basic human objectives of Maslow (1943), presented in Section 2.2.6. More weight is given to the objectives at the bottom of the pyramid, but visceral state affects the weighting, as we shall describe.

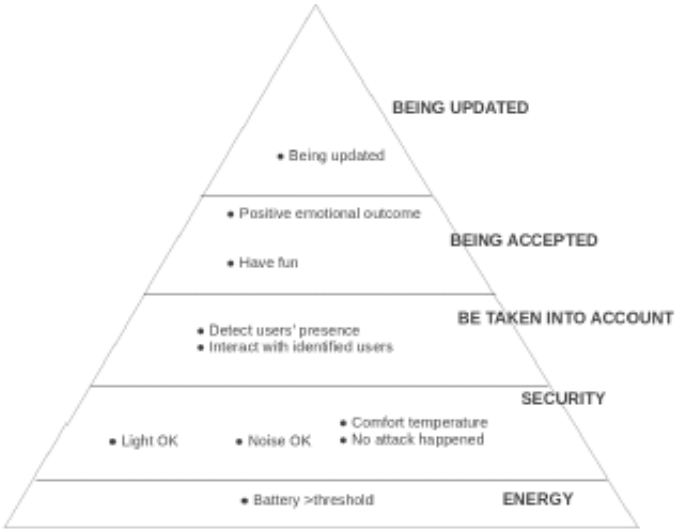


Figure 3.1: The objectives pyramid (Esteban 2012)

The utilities depend on the actions and the environment. The rules for inferring actions and interpreting the environment, as well as the forecasting models, model averaging and the utility functions are explained in more detail in Esteban (2012). This work concentrates on the emotional model explained in next section.

3.2. Emotional model

The affect model is constructed as six dynamic variables and two static levels, and it is summarized in Appendix D. The dynamic variables are a four-

dimensional mixed discrete emotions intensity vector $\mathbf{em}_t$, and the two-dimensional vectors; the physical condition vector $\mathbf{pc}_t$ which is a subset of the environmental state, the visceral state $\mathbf{vs}_t$ which combines emotions and the physical condition, the visceral core affect $\mathbf{ca}_t^{vs}$, the mood vector $\mathbf{md}_t$ and the attitude vector $\mathbf{at}_t$. The static levels are the five factor personality vector **per** and the two-dimensional mood baseline **mb**.

The emotions are a subset of basic emotions discussed in Section 2.2.1, the physical condition is a state vector observed with sensors and simulated for pain, Loewenstein (1996) provides a background for the visceral factors, the core affect and the mood are adapted from Russell (2003), and the attitude (see Section 2.2.3) refers to the dynamic attitude towards the user with whom the interaction takes place. The Big Five model (McCrae & P. T. Costa 1997) is used for personality, and the mood baseline is inspired by Schmidtke & Heller (2004) and Kuppens et al. (2010). It has positive pleasure and arousal values by default and it is affected by Extraversion and Neurotism dimensions of personality. We will describe the variables in the following Sections. $\mathbf{em}_t$, $\mathbf{pc}_t$, $\mathbf{vs}_t$, **per** $\in [0, 1]$ and $\mathbf{ca}_t^{vs}$, $\mathbf{md}_t$, $\mathbf{at}_t$, **mb** $\in [-1, 1]$. The dimensions of the variables are

$$
\begin{aligned}
\mathbf{em}_t &= (Happiness, Anger, Sadness, Fear) \\
\mathbf{pc}_t &= (Battery\ deficiency, Pain) \subset \boldsymbol{e}_t \\
\mathbf{vs}_t &= \mathbf{em}_t \cup \mathbf{pc}_t \\
\mathbf{ca}_t^{vs} &= (Pleasure\ (visceral), Activation\ (visceral)) \\
\mathbf{md}_t &= (Pleasure\ (mood), Activation\ (mood)) \\
\mathbf{at}_t &= (Pleasure\ (attitude), Activation\ (attitude)) \\
\mathbf{per} &= (Openness, Concentiousness, Extraversion, Agreeableness, Neurotism) \\
\mathbf{mb} &= (Pleasure\ (baseline), Activation\ (baseline))\,.
\end{aligned}
$$

3.2.1. Mood

Mood can be understood as a weighted, discounted integral of emotions because it captures the feelings left by past emotions, but focuses on the most recent ones. It moves towards the baseline which is mostly determined by personality. Also the physical condition and the attitude affect mood. It is necessary to transform the visceral state into a two-dimensional mood rating. Temperature, noise or light are not considered, but they could be added too. A core affect mapping $\mathbf{M}_{ca}(i)$ gives the (Pleasure, Activation) coordinates for

each state i. The core affect mappings are estimated from Figure 2.4 and they are shown below in Table 3.2. Only the arousal dimension of mood is used for the physical condition. The core affect of the current visceral state is a value weighted average

$$\mathbf{ca}_t^{vs} = \frac{\sum_{vs \in \mathbf{vs}_t} w_{vs} \cdot vs_t \cdot (vs_t \cdot \mathbf{M}_{ca}(vs)) \ + w_{at}\mathbf{at}_t}{\sum_{vs \in \mathbf{vs}_t} w_{vs} \cdot vs_t \ + w_{at}}, \tag{3.11}$$

where w_{vs} and w_{at} are the core affect weights for the visceral factors and the attitude, respectively.

The mappings are weighted with the visceral intensities in addition to weighting with a normalized product of assigned weights and intensities. This allows us to both scale the area in which the core affect is and weigh the factors according to intensities. The weights can be modified to focus more on areas that are lacking important factors. For example, there is only one positive emotion and three negative ones, so happiness could be weighted more than the other emotions. See Figure 3.2. The dynamic mood model is

$$\mathbf{md}_t = \delta_t^{md}\mathbf{md}_{t-1} + \delta_t^{vs}\mathbf{ca}_t^{vs} + \delta_t^{mb}\mathbf{mb}, \tag{3.12}$$

where δ_t^{md}, δ_t^{vs} and δ_t^{mb} are the normalized weights for the previous mood, visceral core affect and baseline, respectively.

The weights for mood dynamics are determined by visceral values vs_t^i and thresholds $\overline{vs_{md}^i}$, the distance between the previous mood and the visceral core affect $\overline{D_{md,ca}}$, and time since last mood model update T_{mdu}. The ratio between previous mood and the visceral core affect is directly proportional to the distance because mood congruency prevents very different emotions from affecting the mood a lot, and inversely proportional to the update time because if the model is updated often, the mood movements should not be as

i	$\mathbf{M}_{ca}(i)$: (Pleasure, Arousal)
Happiness	(0.98,0.19)
Sadness	(-0.96,-0.27)
Anger	(-0.87,0.5)
Fear	(-0.63,0.78)
Battery deficiency	(0,-1)
Pain	(0,1)

Table 3.2: Core Affect mappings

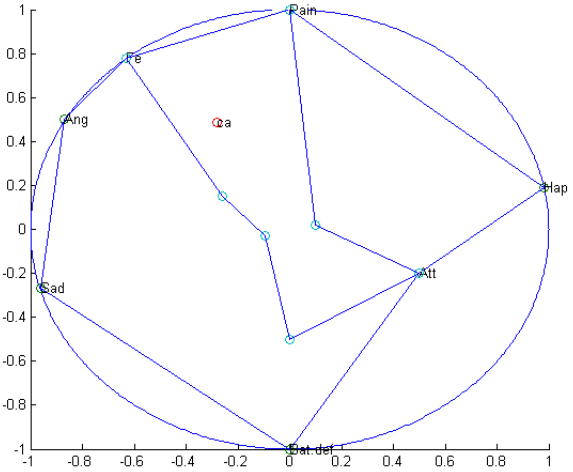


Figure 3.2: An example of a core affect calculation with intensities $\mathbf{vs}_t = [0.1, 0.1, 0.3, 1, 0.5, 1]$, weights $\mathbf{w}_{vs} = [2, 1, 1, 2, 0.5, 0.5]$ and attitude $\mathbf{at}_t = (0.5, -0.2)$ with a weight $w_{at} = 0.5$

big. If the emotional core affect weight is zero, it is replaced with the mood baseline as shown below. The ratio between the emotional core affect and the mood baseline is proportional to the maximum difference between a visceral value and the mood change activation threshold of that factor because when no emotions are very intense (none or minor external events), they do not affect the mood which then starts to go towards the baseline due to internal events. The parameters are

$$
\begin{aligned}
D_{x,y} &= ||x - y||_2 \\
I_{dmax} &= \max(0, D_{at,0} - \underline{at_{md}}, \max_i(vs_t^i - \underline{vs_{md}^i})) \\
\frac{\delta_t^{md}}{\delta_t^{vs}} &= A_{mdvs}\frac{D_{md(t-1),ca}}{T_{mdu}}, \text{ if } I_{dmax} > 0 \qquad (3.13) \\
\frac{\delta_t^{md}}{\delta_t^{mb}} &= A_{mdmb}\frac{D_{md(t-1),mb}}{T_{mdu}}, \text{ if } I_{dmax} = 0 \qquad (3.14) \\
\frac{\delta_t^{vs}}{\delta_t^{mb}} &= A_{vsmb}I_{dmax} \qquad (3.15) \\
\delta_t^{md} + \delta_t^{vs} + \delta_t^{mb} &= 1 \\
\delta_t^{md}, \delta_t^{vs}, \delta_t^{mb} &\geq 0,
\end{aligned}
$$

where $D_{x,y}$ is the Euclidean 2-norm distance between x and y, I_{dmax} is the maximum intensity difference, T_{mdu} is the mood model update time, $\underline{at_{md}}$ is the threshold of attitude and $\underline{vs^i_{md}}$ is the threshold of visceral factor i for influence on mood, δ_t^{md} is the previous mood weight, δ_t^{vs} is the visceral core affect weight, δ_t^{mb} is the mood baseline weight, A_{ij} is the ratio strength parameter for variables i and j.

The normalization of weights ensures that the new mood is inside the core affect circumplex, as long as the initial mood, the mood baseline and the visceral core affect are inside it. The model update time should be small enough to be able to capture the effects of new events that cause emotions. An example of a mood change can be seen in Figure 3.3.

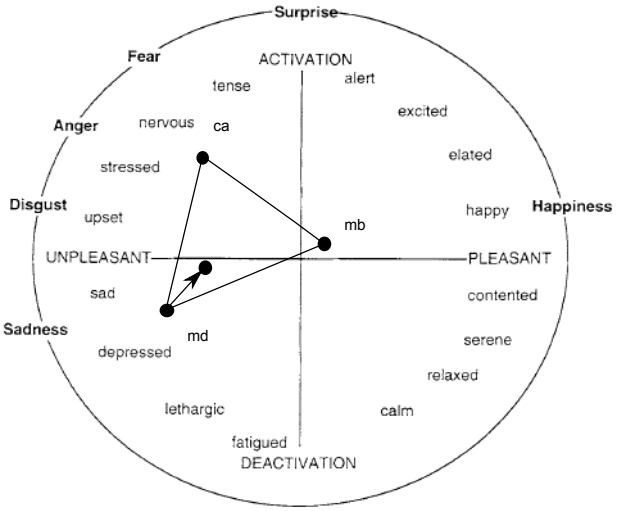


Figure 3.3: An example of a mood change

3.2.2. Emotions

Emotions are generated by action events as well as the physical condition. Memories of past events affect judgement and expectetations of future events are shaped by them. The physical condition is the source of drives that also influence behavior, or action selections, directly. For the robot, the physical condition is comprised of simulated pain, temperature and battery level. Pain occurs when the robot is being hit and elicits anger. A high amount of pain causes the robot to ask for help and attempt to change its place. The external temperature and the battery level are measured with sensors. A bad temperature and low battery level elicit sadness. Events are appraised

according to Roseman et al. (1990). Motive-consistency is measured with utility and expected utility, appetitive or aversive motivation is chosen according to that utility, certainty is based on a standard deviation risk measure, and agency depends on whether the appraisal takes place after the agent's (self/circumstances) or the user's (other/circumstances) turn. Other-agency is only observed in certain events which limits the scope of anger to those events.

Gratch et al. (2009) reviewed different utility-based discrete emotion intensity models. The models were divided in expected utility, expectation-change, threshold, additive and hybrid models. Expected utility models are of form $I_{em} = a \cdot U^p \cdot P^q + b$, expectation-change models $I_{em} = a \cdot U^p \cdot \Delta P^q + b$, threshold models $I_{em} = a \cdot U^p \cdot P^q + b$, if $c \leq P \leq d$, additive models $I_{em} = a \cdot U^p + b \cdot P^q$, and hybrid models a mixture of these types. U is the subjective utility and P is the subjective probability of winning, and ΔP is a probability change within a period. They tested the different models by experimenting with games of Battleship where the subjects could win or lose money. The game was played against a confederate who saw the subjects' positions and could manipulate the game to alter perceptions of winning. The subjects thought that they were playing against each other. The subjective evaluations of winning probabilities, utilities and emotional intensities were recorded at three points. Strongest support was received for the expected utility model, and all the other models except the additive were judged insignificant.

The problem in the article was that framing the game as winning or losing was not effective, probably because the subjects anyway felt they were gaining something (the experiment fee). Negative outcomes then were not considered properly, as a highly probable small utility scenario elicited less sadness than a less probable small utility scenario. That is, even scenarios with bad utility are wanted to be probable, and only probabilities and utilities related to positive goal attainment are modeled.

In our model, there are several possible outcomes through a user reaction to the robot's behavior. When an action a_t is selected, there are valid outcome estimates $\hat{u_t} = u(a_t, b_t, \mathbf{e}_t)$ with probabilities $\hat{p_t} = p(\mathbf{e}_t, b_t \mid a_t, (\mathbf{e}_{t-1}, a_{t-1}, b_{t-1}), (\mathbf{e}_{t-2}, a_{t-2}, b_{t-2}))$. The expected outcome utility and respective modified standard deviation risk measures are

$$\hat{U}_t = \sum_{\mathbf{e}_t, b_t} \hat{p}_t \cdot \hat{u}_t \tag{3.16}$$

$$\sigma^+_{\hat{U}_t} = \sqrt{\sum_{\mathbf{e}_t, b_t : \hat{u}_t > \hat{U}_t} \hat{p}_t \cdot (\hat{u}_t - \hat{U}_t)^2} \tag{3.17}$$

$$\sigma^-_{\hat{U}_t} = \sqrt{\sum_{\mathbf{e}_t, b_t : \hat{u}_t < \hat{U}_t} \hat{p}_t \cdot (\hat{u}_t - \hat{U}_t)^2}. \tag{3.18}$$

These measures can be used to calculate emotional content for the motive-inconsistent emotions fear (aversive: low utility and low positive variation / appetitive: high utility and high negative variation) and sadness (appetitive: high utility and high negative variation) prior to the user reaction ($t-$). The expected emotional contents for time t are

$$ec^{fear}_{t-} = \begin{cases} a^{fear}_{ec-} \cdot (1 - \hat{U}_t) \cdot (1 - \sigma^+_{\hat{U}_t}) + b^{fear}_{ec-} & \text{if } \hat{U}_t < \hat{U}^{fear}_{low} \\ a^{fear}_{ec-} \cdot \hat{U}_t \cdot \sigma^-_{\hat{U}_t} + b^{fear}_{ec-} & \text{if } \hat{U}_t > \hat{U}^{fear}_{high} \\ 0 & \text{else} \end{cases} \tag{3.19}$$

$$ec^{sadness}_{t-} = \begin{cases} a^{sadness}_{ec-} \cdot \hat{U}_t \cdot \sigma^-_{\hat{U}_t} + b^{sadness}_{ec-} & \text{if } \hat{U}_t > \hat{U}^{sadness}_{high} \\ 0 & \text{else}, \end{cases} \tag{3.20}$$

where ec^{em}_{t-} is the expected emotional content of emotion em, and a^{em}_{ec-} and b^{em}_{ec-} are the affine function parameters for it.

When the user reacts with action b_t, the obtained utility $u_t = u(a_t, b_t, \mathbf{e}_t)$ can be compared to the expectation, and a disappointment measure, similar to the Decision Affect Theory (DAT) of Mellers et al. (1997), is constructed. The measure is

$$dp_t = (1 - \hat{p}_t) \cdot df(u_t - \hat{U}_t) \tag{3.21}$$

$$df(x) = \begin{cases} x^{kp} & \text{if } x \geq 0 \\ -\left|x^{kn}\right| & \text{if } x < 0, \end{cases} \tag{3.22}$$

where dp_t is the disappointment measure, $\hat{p}_t$ is the latest probability estimate for the occurred event with the utility u_t and $df()$ is the disappointment function that has parameters kp and kn for positive and negative differences, respectively.

Now emotional content for happiness (high utility or positive surprise), anger (low utility or negative surprise), sadness (low utility) and fear (under attack) can be calculated as follows

$$ec_t^{happiness} = a_{ec}^{happiness} \cdot (u_t + dp_t) + b_{ec}^{happiness} \tag{3.23}$$

$$ec_t^{anger} = a_{ec}^{anger} \cdot (1 - (u_t + dp_t)) + b_{ec}^{anger} \tag{3.24}$$

$$ec_t^{sadness} = \begin{cases} a_{ec}^{sadness} \cdot (1 - u_t) + b_{ec}^{sadness} & \text{if } \hat{U}_t > \hat{U}_{high}^{sadness} \\ 0 & \text{else} \end{cases} \tag{3.25}$$

$$ec_t^{fear} = \min\ (1, ec_{t-1}^{fear} + 0.5) \text{ if } b_t = attack \tag{3.26}$$

$$ec_t^i = \max\left(0, \min\left(1, ec_t^i\right)\right) \Rightarrow ec_t^i \in [0, 1]\ \forall i,$$

where ec_t^{em} is the emotional content of emotion em, a_{ec-}^{em} and b_{ec-}^{em} are the affine function parameters for it, and $\hat{U}_{high}^{sadness}$ is the utility threshold for experiencing sadness.

Emotional content causes emotion intensities depending on the mood and personality. The TAME model (Moshkina 2006) introduces an intensity function, shown in Equation 1.3. We use core affect distances between the mood and the emotions, as well as the personality-emotion mappings $M_{pe}(i, j)$ shown in Table 3.3, to form a similar, dynamic growth function

Emotion \ Personality	O	C	E	A	N
Happiness	1	0	1	1	0
Sadness	1	0	0	0	1
Anger	1	0	0	-1	1
Fear	1	0	0	0	1

Table 3.3: Personality effects on emotions $M_{pe}(i, j)$ (Moshkina 2011)

$$em_t^i = \frac{pv^i}{1 + ap^i \cdot e^{-gr^i \cdot (ec_t^i - 0.5)}} \tag{3.27}$$

$$pv^i = \min\left(1,\ 0.7 + \frac{0.5}{N_{M_{pe} \neq 0}} \sum_{\mathbf{per}} per_j \cdot M_{pe}(i,j)\right) \tag{3.28}$$

$$gr^i = 4 + \frac{9}{N_{M_{pe} \neq 0}} \sum_{\mathbf{per}} per_j \cdot M_{pe}(i,j) \tag{3.29}$$

$$ap_t^i = 6 + 10 D_{\mathbf{M}_{ca}(em^i), \mathbf{md}_t} - \frac{5}{N_{M_{pe} \neq 0}} \sum_{\mathbf{per}} per_j \cdot M_{pe}(i,j) \tag{3.30}$$

where em_t^i is emotion intensity, $N_{M_{pe} \neq 0}$ is the amount of nonzero personality-emotion mappings for emotion i, ec_t^i is the emotional content, ap_t^i controls the activation point, gr^i controls the maximum slope, and pv^i is the peak value.

With a personality $\mathbf{per} = (1, 0, 0, 1, 0)$ and a mood $\mathbf{md}_t = (-0.98, -0.19)$, the parameters for happiness would be $pv^{happiness} = \min(1,\ 0.7 + \frac{0.5}{3} \cdot (1 \cdot 1 + 0 \cdot 1 + 1 \cdot 1)) = 1$, $gr^{happiness} = 4 + \frac{9}{3} \cdot (1 \cdot 1 + 0 \cdot 1 + 1 \cdot 1) = 10$ and $ap_t^{happiness} = 6 + 10 \cdot 2 - \frac{5}{3} \cdot (1 \cdot 1 + 0 \cdot 1 + 1 \cdot 1) = 22\frac{2}{3}$. Examples of intensity functions can be seen in Figure 3.4. Emotions are instantaneous and they are evaluated according to action events, i.e. after the robot's action a_t and after the user's action b_t. The jumps can be smoothened with additional dynamics if the situation requires it.

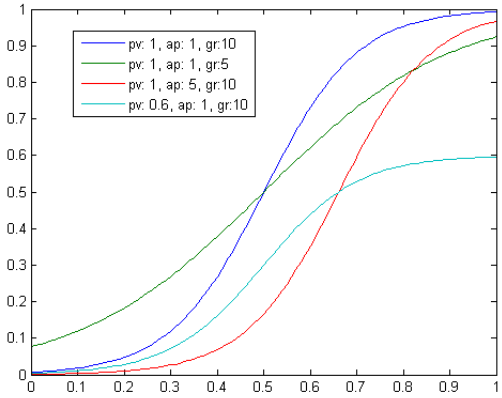


Figure 3.4: Emotion intensity function examples

3.2.3. Attitude

The user-specific attitude represents the emotional aspect of the relationship. A familiarity measure, adapted from Kirby et al. (2010), is used in updating perceptions of existing users and in memory allocation for new users. The more familiar a person is, the less weight is given to new data and the more likely that the user is remembered. Also the Openness dimension of personality affects the updating. A default attitude profile is used for unrecognized users before there is enough data to develop an attitude towards them. The default profile is based on first interactions with unrecognized people. Attitude is formed through conditioning, i.e. the emotions that are associated with interacting with the user, as follows

$$fam_t^{B_u} = \frac{1}{2}\left[1 + \frac{1}{10}\min(10, int_{tot}(B_u)) - \frac{1}{30}\min(30, int_{prev}(B_u))\right] \forall B_u \in \mathcal{B}_u \tag{3.31}$$

$$fam_t^{cur} \leftarrow \min_{B_u} fam_t^{B_u} \text{ if } fam_t^{cur} > \min_{B_u} fam_t^{B_u} \tag{3.32}$$

$$\mathbf{at}_t(def) = w_t^{at,old} \cdot \mathbf{at}_{t-1}(def) + w_t^{at,new} \cdot \mathbf{ca}_t^{vs} \text{ if } int_{tot}(cur) < 2 \tag{3.33}$$

$$\mathbf{at}_t(B_u) = w_t^{at,old} \cdot \mathbf{at}_{t-1}(B_u) + w_t^{at,new} \cdot \mathbf{ca}_t^{vs} \tag{3.34}$$

$$\frac{w_t^{at,old}}{w_t^{at,new}} = A_{att}\frac{0.8 \cdot fam_t^{B_u} - 0.2 \cdot per_O}{T_{mdu}} \tag{3.35}$$

$$w_t^{at,old} + w_t^{at,new} = 1,$$

where $fam_t^{B_u}$ is the familiarity measure, $int_{tot}(B_u)$ and $int_{prev}(B_u)$ are the total (hours) time of and the time since the previous (days) interaction with the user B_u, cur refers to the most probable current user, $\mathbf{at}_t(def)$ and $\mathbf{at}_t(B_u)$ are the default and user-specific attitudes, $w_t^{at,old}$ and $w_t^{at,new}$ are the weights for old and new attitude data, and A_{att} is the attitude ratio strength parameter for old and new attitudes. The visceral core affect $\mathbf{ca}_t^{vs}$ is calculated with Equation 3.11. The arrow ($\leftarrow$) in Equation 3.32 indicates a memory reallocation.

3.2.4. Fear and pain

Bolles & Fanselow (1980) introduced a recuperative model of fear and pain. In a critical situation, e.g. under attack, fear is elicited and pain is inhibited until the subject has successfully dealt with the threat and can safely take

care of the wounds. In this framework, the only user action to cause pain is $b_t = attack$.

$$pain_t = \begin{cases} \min\ (1, \frac{A_{pn}\delta_{pd}}{T_{mdu}} \cdot pain_{t-1} + 0.5) & \text{if } b_t = attack \\ \frac{A_{pn}\delta_{pd}}{T_{mdu}} \cdot pain_{t-1} & \text{else} \end{cases}, \tag{3.36}$$

where A_{pn} is the pain normalization factor and δ_{pd} is the pain decay parameter. The inhibition effect is taken into account in the decision making module, where fear and pain affect the robot's action probabilities.

3.2.5. Heuristic action evaluation

In some emotional states, specifically happiness, anger and disgust as shown in Table 2.4, analytic processing is replaced with, or overweighted by, heuristic processing. An emotion-based measure is then used to select actions. We use Q-learning, as in Cattinelli et al. (2008), to assign a measure for actions.

$$Q(st_{t-1}, a_{t-1}) \leftarrow_t Q(st_{t-1}, a_{t-1}) + \alpha \left[R_t + \gamma \max_{a_t} Q(st_t, a_t) - Q(st_{t-1}, a_{t-1}), \right] \tag{3.37}$$

where $Q(st_t, a_t)$ is the Q-value for action a_t in state st_t, α is the learning rate, R_t is the reward achieved after performing a_{t-1} in state st_{t-1}, and γ is the discount factor.

The Q-values are updated when the rewards for previous actions are observed. The rewards can be based on utility or emotional valence of the consequences of the user's reaction to the agent's action. In this case, the valence component of visceral core affect is used.

$$R_t = (1\ 0) \cdot \mathbf{ca}_t^{vs} \tag{3.38}$$

The states are discrete, so in addition to emotional intensities, the physical condition must be categorized. The categories are shown in Table 3.4. Each state is the combination of the emotion with the highest intensity and the physical condition, which makes up $4 \cdot 2 \cdot 2 \cdot 2 = 32$ possible states. For example, one possible state is $st_t = (high\ fear, high\ pain, low\ bat.def)$.

Visceral factor	Intervals for low and high
Emotions	$[0, \frac{1}{2}], (\frac{1}{2}, 1]$
Physical condition	$[0, \frac{1}{2}], (\frac{1}{2}, 1]$

Table 3.4: Visceral state categories

3.2.6. Effects on the Decision Model

Emotions have effects on how risk is evaluated. When the probabilities of different scenarios are calculated using forecasting models in decision making and emotion generation, they are modified according to the current visceral state. Table 2.4 summarizes the effects. First, let us define a decision weight function as in Gonzalez & Wu (1999).

$$w(p) = \frac{\delta_p p^{\gamma_p}}{\delta_p p^{\gamma_p} + (1-p)^{\gamma_p}}, \tag{3.39}$$

where the default parameter values are 0.44 for attractiveness (δ_p) and 0.77 for discriminality (γ_p).

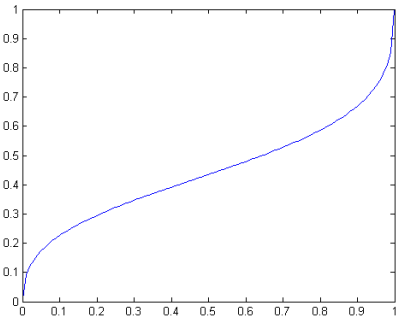


Figure 3.5: The decision weight function

Now, the actual probability that is used in decision making and emotion dynamics is further biased with risk attitude. For optimistic risk perception, the higher the utility and optimism, the higher its probability is perceived. Both emotions and mood affect risk attitude.

$$ra_t = \frac{1}{N_{em}} \left(\sum_{em^i \in opt} em_t^i - \sum_{em^i \in pes} em_t^i \right) + (1\ 0) \cdot \mathbf{md}_t \tag{3.40}$$

$$p_t = (1 + ra_t \cdot (u_t - 0.5)) \cdot w(p), \tag{3.41}$$

where ra_t is the momentary risk attitude, N_{em} is the amount of emotions, opt refers to the emotions with an optimistic and pes to the emotions with a pessimistic risk perception, and p_t is the perceived probability.

Emotions also affect the processing strategy. The more intense analytic emotions are, the more the expected utility calculations are used. Openness can decrease the confidence rate which flattens the action probabilities.

$$ah_t = \frac{1}{N_{em}} \left(\sum_{em^i \in ana} em_t^i - \sum_{em^i \in heu} em_t^i \right) \tag{3.42}$$

$$cf = 1 + 3 \cdot (0.8 - per_O) \tag{3.43}$$

$$P(a_t) \propto \left(A_{ps} \cdot (1 + ah_t) \cdot \psi(a_t) + Q(st_t, a_t) \right)^{cf}, \tag{3.44}$$

where ah_t is the momentary analyticity, ana refers to the emotions with an analytic and heu to the emotions with a heuristic processing strategy, per_O is the openness value of personality, $P(a_t)$ is the probability of choosing action a_t and A_{ps} is the processing strategy normalization factor making expected utilities and Q-values comparable. Compare to Equation 3.10.

The visceral state alters the weights for objectives. Positive emotions promote higher order needs, fear and pain increase safety need, and battery deficiency increases energy need.

$$W_t^1 = 10 - 4R_t + 5\ bat.def._t \tag{3.45}$$

$$W_t^2 = 8 - 2R_t + 5 \cdot (pain_t + fear_t) \tag{3.46}$$

$$W_t^3 = 4 + R_t \tag{3.47}$$

$$W_t^4 = 2 + 2R_t \tag{3.48}$$

$$W_t^5 = 1 + 3R_t \tag{3.49}$$

$$w_t^{ob,i} = \frac{W_t^i}{\sum_i W_t^i}, \tag{3.50}$$

where W_t^i and $w_t^{ob,i}$ are the weights and normalized weights of the objectives of energy, security, being taken into account, being accepted and being updated, and R_t is the pleasure component of visceral core affect shown in Equation 3.38.

Sometimes visceral factors can have a direct effect on behavior. A high amount of fear increases probabilities for *cry, alert* and *warn* actions. Otherwise, high pain will cause the *ask for help* and high battery deficiency the *ask for charging* action.

3.3. Robot features

The robot used in the model development can be seen in Figure 3.6. It has a microprocessor, a LINUX-based operating system and sensors for temperature, inclination, touch, light and strength. Its camera and audio system enable user recognition and identifying conversation topics. The audio system can also be used for estimating emotional content in the interaction. The robot's sensors, computational capacity and physical composition always place constraints on the input and output of the model. (García, Pallardó, Insua, Moreno & Redchuk 2012)

Figure 3.6: AISoy1 Robot (AISoy Robotics S.L. 2012)

3.4. Further research

Emotional models can never be perfect, not least because emotion is such a complex issue that psychological and neurological research fails to provide

coherent theories on it. Furthermore, each individual has different emotional dynamics due to genes, personal history, culture and environment. In spite of this, we can make the models more sophisticated and place focus on some area depending on the application domain. Often it is not useful to replicate exact human emotions for a robot, but rather to copy selected features to enhance interaction.

In our model, the visceral state is only a subset of all actual factors. There is no sex drive or disgust. The visceral core affect is biased and there should be a complementary offset guiding the mood in addition to the adjusted weighting. The model should be expanded by adding new emotions and other visceral factors when there is satisfactory research on their dynamics and effects on decision making. Cultural aspects are ignored even though they may be very important. For example, people in the United States prefer anger to fear whereas the Machiguenga Indians in the Peruvian Amazon prefer fear to anger (Lerner & Keltner 2001). Emotion expression was not covered in this work, but it is a very integral part of emotional communication and it should be implemented in the robot. See Álvarez, Galán, Matía, Rodríguez-Losada & Jiménez (2010) and Kirby et al. (2010) for examples.

Emotion regulation affects the emotional system in five stages: selecting the situation, modifying the situation, deploying attention, changing cognition and modulating the responses. In contrast to mood regulation, emotion regulation targets specific emotions and associated responses (Gross 1998). Avoiding sad states or expressing more anger could be useful in manipulating the environment.

In social situations, a framework is needed for group engagement and roles. When to initiate or conclude communication, how to understand the roles and statuses of the participants and how to model social emotions e.g. shame. Keltner & Haidt (1999) noted that emotions have diverse social functions. They increase group solidarity, act as behavioral cues for children and exhibit social status. Emotional expression generates emotions, for example anger causes fear. Emotion recognition would be an important addition in the model.

Estimating the user's utility in various states would enable e.g. gratification and revenge. It is quite difficult because the user's objectives need to be assessed and actions linked to relevant goals. A model would have to be developed for each user and the parameters evaluated using interaction data. Perhaps pattern recognition would be useful in this task.

Humans learn new actions by observing other people and by studying. A robot might learn new actions by communicating with other robots or by examining user actions, although an advanced module for recognizing objects and

entities would be mandatory. Also in this case, evaluating goal relatedness of actions is vital. In a goal-oriented framework, the selection of goals, task management and causal deduction are additional requirements. Ortony et al. (1988) provide a popular model.

Further context identification and learning would be advantageous. For example, knowing whether the robot is inside or outside would help in inferring whether the user or the circumstances are responsible for temperature changes. The agency and power, and motive-consistency in a goal-oriented framework, dimensions of event appraisal (Roseman et al. 1990) are possible to estimate only if context evaluation and inference modules exist.

Finally, development stages for emotion dynamics create more diverse and adaptive results. Álvarez et al. (2010) divided the evolution of the emotional system in three stages: the infant, the youth and the adult stage. In the infant stage, the robot modified its emotional reactions based on external input, e.g. a sentence "an insult should not make you so sad". In the youth stage, the robot develops its emotional behaviour so that learns patterns to maximize its happiness. In the adult stage, it can adapt to its environment as a means of coping, and it may change its values accordingly.

Chapter 4

Conclusions

In this thesis, we have introduced an emotional model for a social robot based on event appraisal that acts together with a decision making model guiding the robot's behavior. The purpose of the work was to enhance the framework of Esteban (2012) by adding a sophisticated humanlike emotional system, which includes multilevel affect, comprised most notably of mood and discrete emotions, personalities, relationships as attitudes, and advanced emotional dynamics using event appraisal, mood congruency and physical condition. The emotional system should make the robot smoother and more believable in social contexts.

Choosing a framework for the emotional system was very challenging because there are a lot of competing theories on the different parts and components of affect. There is even uncertainty on what emotions are (Russell 2003). There are debates on whether there are some basic emotions (Ekman & Friesen 1971, Ortony & Turner 1990, Levenson et al. 1992, Izard 1992, Panksepp 1992, Turner & Ortony 1992), what the dimensions of emotions are (Watson et al. 1988, Russell & Barrett 1999, Larsen & McGraw 2001), how emotions evolve (Zajonc 1984, Lazarus 1984, Frijda et al. 1989, Roseman et al. 1990), and what their effects on decisions are (Leone et al. 2005, Forgas 1995).

We succeeded in using several related theories to create a unified emotional system. It is centered around a two-dimensional core affect. Mood integrates the effects of momentary visceral states and a personality-related baseline, and it is described on the core affect plane. The visceral state, which involves emotions and the physical state, is mapped to the core affect plane using intensities as weights. The emotional contents of events are generated using utilities, probabilities and related risk measures as event appraisal dimensions. A disappointment measure is used for comparing expected and

occurred events. The emotional content is transformed into emotional intensities through a dynamic activation function which controls mood and trait congruency of emotions.

The emotional system not only takes input from the decision making module, but also affects its operation. The different risk perceptions of emotions form a momentary risk attitudes based on momentary intensities and mood. The risk attitude then shapes perceived probabilities, or decision weights, of events. Emotions differ in processing strategies as well. They can provoke analytic or heuristic thinking, and for a heuristic-focused emotional state, action selection is influenced more by the values provided by Q-learning than by the utilities and probabilities acquired from the forecasting models. Finally, emotions affect the weighting of objectives. Negative emotions set the focus on lower order needs, energy and security, whereas positive emotions promote higher aspirations. Overall, the research objectives were well achieved.

The introduced framework is valuable, because it combines the emotional and decision making systems and enables affective behavior which is crucial to the sociable humans. The core affect measure of discrete emotions and other visceral factors links the ambiguous, continuous mood with more easily identifiable affective states. Having an easily expandable multilevel affect allows us to capture the various effects of the complex emotional system on decision making. It is especially important to understand that emotions operate and influence decisions on different levels which are interconnected. Emotions can be linear in some parts of the system, but, on the whole, they are very much non-linear and chaotic.

Further research is needed to estimate the parameters and test the validity of the model. Simulation is often impossible, because the events depend on the model-based robot behavior *and* the human users, whose reactions cannot be predicted before there is interaction data. When parameters are changed, this process has to be restarted. Gathering a data set on user reactions and environment evolution in different settings could be useful in testing the various parts of the model and configuring initial parameters. The users decide whether the robot is functional in its domain.

Bibliography

AISoy Robotics S.L. (2012), 'AISoy Robotics'.
URL: http://www.aisoy.es/

Álvarez, M., Galán, R., Matía, F., Rodríguez-Losada, D. & Jiménez, A. (2010), 'An emotional model for a guide robot', *IEEE Transactions on Systems, Man, and Cybernetics* **40**(5), 982–992.

Anderson, C. & Galinsky, A. D. (2006), 'Power, optimism, and risk-taking', *European Journal of Social Psychology* **36**(4), 511–536.

Batson, C. D. & Ahmad, N. (2001), 'Empathy-induced altruism in a prisoner's dilemma II: what if the target of empathy has defected?', *European Journal of Social Psychology* **31**(1), 25–36.

Bechara, A., Damasio, H. & Damasio, A. R. (2000), 'Emotion, decision making and the orbitofrontal cortex', *Cerebral Cortex* **10**(3), 295–307.

Beedie, C. J., Terry, P. C. & Lane, A. M. (2005), 'Distinctions between emotion and mood', *Cognition and Emotion* **19**(6), 847–878.

Bellman, R. (1957), *Dynamic Programming*, Princeton University Press, Princeton, NJ.

Berg, J., Dickhaut, J. & McCabe, K. (1995), 'Trust, reciprocity, and social history', *Games and Economic Behavior* **10**(1), 122–142.

Bolles, R. C. & Fanselow, M. S. (1980), 'A perceptual defensive recuperative model of fear and pain', *Behavioral and Brain Sciences* **3**(2), 291–301.

Busemeyer, J. R., Dimperio, E. & Jessup, R. K. (2007), Integrating emotional processes into decision making models, *in* W. Gray, ed., 'Integrated models of cognitive systems', Oxford University Press.

Camerer, C. & Thaler, R. H. (1995), 'Ultimatums, dictators and manners', *Journal of Economic Perspectives* **9**(2), 209–219.

Cattinelli, I., Goldwurm, M. & Borghese, N. A. (2008), 'Interacting with an artificial partner: modeling the role of emotional aspects', *Biological Cybernautics* **99**(6), 473–489.

Clemen, R. (1996), *Making hard decisions: An introduction to decision analysis*, Duxbury, Pacific Grove, CA.

Clore, G. L. & Schnall, S. (2005), The influence of affect on attitude, *in* D. Albarracín, B. T. Johnson & M. P. Zanna, eds, 'Handbook of attitudes', Lawrence Erlbaum Associates Publishers.

Cohen, J. D. (2005), 'The vulcanization of the human brain: A neural perspective on interactions between cognition and emotion', *Journal of Economic Perspectives* **19**(4), 3–24.

Damasio, A. (1994), *Descartes' Error: Emotion, Reason, and the Human Brain*, Putnam.

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L. B., Parvizi, J. & Hichwa, R. D. (2000), 'Subcortical and cortical brain activity during the feeling of self-generated emotions', *Nature Neuroscience* **3**(10), 1049–1056.

Damasio, H. & Grabowski, T. (1994), 'The return of Phineas Gage: Clues about the brain from the skull of a famous patient', *Science* **264**(5162), 1102–1105.

de Vries, M., Holland, R. W. & Witteman, C. L. M. (2008), 'Fitting decisions: Mood and intuitive versus deliberative decision strategies', *Cognition and Emotion* **22**(5), 931–943.

Delgado, M. R., Frank, R. H. & Phelps, E. A. (2005), 'Perceptions of moral character modulate the neural systems of reward during the trust game', *Nature Neuroscience* **8**, 1611–1618.

Diener, E., Suh, E. M., Lucas, R. E. & Smith, H. L. (1999), 'Subjective well-being: Three decades of progress', *Psychological Bulletin* **125**(2), 276–302.

Eid, M.; Diener, E. (1999), 'Intraindividual variability in affect: Reliability, validity, and personality correlates', *Journal ol Personality and Social Psychology* **76**(4), 662–676.

Ekman, P. (1992), 'Are there basic emotions', *Psychological Review* **99**(3), 550–553.

Ekman, P. & Friesen, W. V. (1971), 'Constants across cultures in the face and emotion', *Journal ol Personality and Social Psychology* **17**(2), 124–129.

El-Nasr, M. S., Yen, J. & Ioerger, T. R. (2000), 'FLAME – Fuzzy logic adaptive model of emotions', *Autonomous Agents and Multi-Agent Systems* **3**(3), 219–257.

Elliott, C. (1992), The affective reasoner: A process model of emotions in a multi-agent system, PhD thesis, Northwestern University Institute for the Learning Sciences, Northwestern, IL.

Engel, C. (2010), Dictator games: A meta study, Technical report, Max Planck Institute.

Engle-Warnick, J. & Slonim, R. L. (2004), 'The evolution of strategies in a repeated trust game', *Journal of Economic Behavior & Organization* **55**(4), 553–573.

Esteban, P. G. (2012), Modeling and implementing an emotional based decision agent. To appear in proceedings of XXXIII Congreso Nacional de Estadística e Investigación Operativa.

Fessler, D. M. T., Pillsworth, E. G. & Flamson, T. J. (2004), 'Angry men and disgusted women: An evolutionary approach to the influence of emotions on risk taking', *Organizational Behavior and Human Decision Processes* **95**(1), 107–123.

Forgas, J. P. (1995), 'Mood and judgment: The affect infusion model (AIM)', *Psychological Bulletin* **117**(1), 39–66.

Frank, M. J. & Claus, E. D. (2006), 'Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal', *Psychological Review* **113**(2), 300–326.

Frijda, N. H., Kuipers, P. & ter Schure, E. (1989), 'Relations among emotion, appraisal, and emotional action readiness', *Journal of Personality and Social Psychology* **57**(2), 212–228.

García, D., Pallardó, C., Insua, D., Moreno, R. & Redchuk, A. (2012), 'Aisoy 1: A robot that perceives, feels and makes decisions'.

URL: http://ercim-news.ercim.eu/en84/special/aisoy-1-a-robot-that-perceives-feels-and-makes-decisions

Glimcher, P., Camerer, C., Fehr, E. & Poldrack, R. (2008), *Neuroeconomics: Decision making and the brain*, AcademicPress, New York.

Gonzalez, R. & Wu, G. (1999), 'On the shape of the probability weighting function', *Cognitive Psychology* **38**(1), 129–166.

Gratch, J. & Marsella, S. (2004), 'A domain-independent framework for modeling emotion', *Journal of Cognitive Systems Research* **5**(4), 269–306.

Gratch, J., Marsella, S., Wang, N. & Stankovic, B. (2009), Assessing the validity of appraisal-based models of emotion, *in* 'International Conference on Affective Computing and Intelligent Interaction', ACII2009.

Gross, J. J. (1998), 'The emerging field of emotion regulation: An integrative review', *Review of General Psychology* **2**(3), 271–299.

Gunnthorsdottir, A., Houser, D. & McCabe, K. (2007), 'Disposition, history and contributions in public goods experiments', *Journal of Economic Behavior & Organization* **62**(2), 304–315.

Hagerty, M. R. (1999), 'Testing maslow's hierarchy of needs: National quality-of-life across time', *Social Indicators Research* **46**(3), 249–271.

Heller, W. (1993), 'Neuropsychological mechanisms of individual differences in emotion, personality, and arousal', *Neuropsychology* **7**(4), 476–489.

Hoeting, J. A., Madigan, D., Raftery, A. E. & Volinsky, C. T. (1999), 'Bayesian model averaging: A tutorial', *Statistical Science* **14**(4), 382–417.

Hofstede, G. (1984), 'The cultural relativity of the quality of life concept', *Academy of Management Review* **9**(3), 389–398.

Izard, C. E. (1992), 'Basic emotions, relations among emotions, and emotion-cognition relations', *Psychological Review* **99**(3), 561–565.

Izard, C. E. (1993), 'Four systems for emotion activation: Cognitive and noncognitive processes', *Psychological Review* **100**(1), 68–90.

Izard, C. E. (1994), 'Innate and universal facial expressions: Evidence from developmental and cross-cultural research', *Psychological Bulletin* **115**(2), 288–299.

Janis, I. L. & Leventhal, H. (1967), Human reactions to stress, *in* E. Borgatta & W. Lambert, eds, 'Handbook of personality theory and research', Rand McNally, Chicago.

Johnson-Laird, P. N. & Oatley, K. (1989), 'The language of emotions: An analysis of a semantic field', *Cognition and Emotion* **3**(2), 81–123.

Kaelbling, L. P., Littman, M. L. & Moore, A. W. (1996), 'Reinforcement learning: a survey', *Journal of Artificial Intelligence Research* **4**, 237–285.

Kahneman, D. (2011), *Thinking, Fast and Slow*, Farrar, Straus and Giroux.

Keltner, D. & Haidt, J. (1999), 'Social functions of emotions at four levels of analysis', *Cognition and Emotion* **13**, 505–521.

Kirby, R., Forlizzi, J. & Simmons, R. (2010), 'Affective social robots', *Robotics and Autonomous Systems* **58**(3), 322–332.

Kuppens, R., Oravecz, Z. & Tuerlinckx, F. (2010), 'Feelings change: Accounting for individual differences in the temporal dynamics of affect', *Journal of Personality and Social Psychology* **99**(6), 1042–1060.

Kurzban, R. & Houser, D. (2005), 'Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations', *Proceedings of the National Academy of Sciences of the United States of America* **102**(5), 1803–1807.

Larsen, J. T. & McGraw, A. P. (2001), 'Can people feel happy and sad at the same time?', *Journal oi Personality and Social Psychology* **81**(4), 684–696.

Lazarus, R. S. (1984), 'On the primacy of cognition', *American Psychologist* **39**(2), 124–129.

LeDoux, J. E. (1995), 'Emotion: Clues from the brain', *Annual Review of Psychology* **46**, 209–235.

Leone, L., Perugini, M. & Bagozzi, R. P. (2005), 'Emotions and decision making: Regulatory focus moderates the influence of anticipated emotions on action evaluations', *Cognition and Emotion* **19**(8), 1175–1198.

Lerner, J. S. & Keltner, D. (2000), 'Beyond valence: Toward a model of emotion-specific influences on judgement and choice', *Cognition and Emotion* **14**(4), 473–493.

Lerner, J. S. & Keltner, D. (2001), 'Fear, anger, and risk', *Journal of Personality and Social Psychology* **81**(1), 146–159.

Levenson, R. W., Ekman, P., Heider, K. & Friesen, W. V. (1992), 'Emotion and autonomic nervous system activity in the Minangkabau of West Sumatra', *Journal of Personality and Social Psychology* **62**(6), 972–988.

Loewenstein, G. (1996), 'Out of control: Visceral influences on behavior', *Organizational Behavior and Human Decision Processes* **65**(3), 272–292.

Loewenstein, G. (2000), 'Emotions in economic theory and economic behavior', *The American Economic Review* **90**(2), 426–432.

Loewenstein, G. & Lerner, J. S. (2003), The role of affect in decision making, *in* R. J. Davidson, K. R. Scherer & H. H. Goldsmith, eds, 'Handbook of Affective Sciences', Oxford University Press.

Mamdani, E. H. & Assilian, S. (1975), 'An experiment in linguistic synthesis with a fuzzy logic controller', *International Journal of Man-Machine Studies* **7**(1), 1–13.

Marcus, G. E. & Mackuen, M. B. (1993), 'Anxiety, enthusiasm, and the vote: The emotional underpinnings of learning and involvement during presidential campaigns', *American Political Science Review* **87**(3), 672–685.

Markon, K. E., Krueger, R. F. & Watson, D. (2005), 'Delineating the structure of normal and abnormal personality: An integrative hierarchical approach', *Journal of Personality and Social Psychology* **88**(1), 139–157.

Maslow, A. H. (1943), 'A theory of human motivation', *Psychological Review* **50**(4), 370–396.

McCrae, R. R. & P. T. Costa, J. (1997), 'Personality trait structure as a human universal', *American Psychologist* **52**(5), 509–516.

McDermott, R. (2004), 'The feeling of rationality: The meaning of neuroscientific advances for political science', *Perspectives on Politics* **2**(4), 691–706.

Mellers, B. A., Schwatz, A., Ho, K. & Ritov, I. (1997), 'Decision affect theory: Emotional reactions to the outcomes of risky options', *Psychological Science* **8**(6), 423–429.

Moshkina, L. (2006), An integrative framework for affective agent behavior, *in* 'Proceedings of the International Conference on Intelligent Systems and Control', IASTED.

Moshkina, L. (2011), An Integrative Framework of Time-Varying Affective Robotic Behavior, PhD thesis, Georgia Institute of Technology.

Öhman, A. & Soares, J. J. F. (1998), 'Emotional conditioning to masked stimuli: Expectancies for aversive outcomes following nonrecognized fear-relevant stimuli', *Journal of Experimental Psychology* **127**(1), 69–82.

Olson, M. A. & Fazio, R. H. (2001), 'Implicit attitude formation through classical conditioning', *Psychological Science* **12**(5), 413–417.

Ortony, A., Clore, G. & Collins, A. (1988), *The Cognitive Structure of Emotions*, Cambridge University Press, Cambridge.

Ortony, A. & Turner, T. J. (1990), 'What's basic about basic emotions?', *Psychological Review* **97**(3), 315–331.

Palm, R., Hodgson, M., Blanchard, D. & Lyons, D. (1990), *Earthquake insurance in California*, Westview Press, Boulder, CO.

Panksepp, J. (1992), 'A critical role for "affective neuroscience" in resolving what is basic about basic emotions', *Psychological Review* **99**(3), 554–560.

Pavlov, I. P. (1927), *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*, Oxford University Press, London.

Pham, M. T. (2007), 'Emotion and rationality: A critical review and interpretation of empirical evidence', *Review of General Psychology* **11**(2), 155–178.

Phan, K. L., Wager, T., Taylor, S. F. & Liberzon, I. (2002), 'Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI', *NeuroImage* **16**(2), 331–348.

Picard, R. (1997), *Affective computing*, MIT Press, Cambridge.

Picard, R. W. (2003), 'Affective computing: challenges', *International Journal of Human-Computer Studies* **59**(1), 55–64.

Price, D. D., Barrell, J. E. & Barrell, J. J. (1985), 'A quantitative-experiential analysis of human emotions', *Motivation and Emotion* **9**(1), 19–38.

Raghunathan, R. & Pham, M. T. (1999), 'All negative moods are not equal: Motivational influences of anxiety and sadness on decision making', *Organizational Behavior and Human Decision Processes* **79**(1), 56–77.

Rázuri, J. G., Esteban, P. G. & Ríos Insua, D. (2011), An adversarial risk analysis model for an emotional based decision agent, *in* 'Neural Information Processing System Conference 2011', NIPS2011.

Reisenzein, R. (1994), 'Pleasure-arousal theory and the intensity of emotions', *Journal of Personality and Social Psychology* **67**(3), 525–539.

Rickel, J. & Johnson, W. L. (1999), 'Animated agents for procedural training in virtual reality: Perception, cognition, and motor control', *Applied Artificial Intelligence* **13**(4-5), 343–382.

Ríos Insua, D., Ríos, J. & Banks, D. (2009), 'Adversarial risk analysis', *Journal of the American Statistical Association* **104**(406), 841–854.

Roseman, I. J., Spindel, M. S. & Jose, P. E. (1990), 'Appraisals of emotion-eliciting events: Testing a theory of discrete emotions', *Journal of Personality and Social Psychology* **59**(5), 899–915.

Russell, J. A. (2003), 'Core affect and the psychological construction of emotion', *Psychological Review* **110**(1), 145–172.

Russell, J. A. & Barrett, L. F. (1999), 'Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant.', *Journal of Personality and Social Psychology* **76**(5), 805–819.

Rusting, C. L. (1998), 'Personality, mood, and cognitive processing of emotional information: Three conceptual frameworks', *Psychological Bulletin* **124**(2), 165–196.

Rusting, C. L. & DeHart, T. (2000), 'Retrieving positive memories to regulate negative mood: Consequences for mood-congruent memory', *Journal of Personality and Social Psychology* **78**(4), 737–752.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. (2003), 'The neural basis of economic decision-making in the ultimatum game', *Science* **300**(5626), 1755–1758.

Scherer, K. R. (2000), Psychological models of emotion, *in* J. C. Borod, ed., 'The Neuropsychology of Emotion', Oxford University Press.

Schmidtke, J. I. & Heller, W. (2004), 'Personality, affect and EEG: predicting patterns of regional brain activity related to extraversion and neuroticism', *Personality and Individual Differences* **36**(3), 717–732.

Schwarz, N. (2000), 'Emotion, cognition, and decision making', *Cognition and Emotion* **14**(4), 433–440.

Smith, C. A. & Lazarus, R. (1990), Emotion and adaptation, *in* L. A. Pervin, ed., 'Handbook of Personality: Theory and research', Guilford Press, New York, NY, pp. 609–637.

Tay, L. & Diener, E. (2011), 'Needs and subjective well-being around the world', *Journal of Personality and Social Psychology* **101**(2), 354–365.

Tiedens, L. Z. & Linton, S. (2001), 'Judgment under emotional certainty and uncertainty: The effects of specific emotions on information processing', *Journal of Personality and Social Psychology* **81**(6), 973–988.

Turner, T. J. & Ortony, A. (1992), 'Basic emotions: Can conflicting criteria converge?', *Psychological Review* **99**(3), 566–571.

Tversky, A. & Kahneman, D. (1992), 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and Uncertainty* **5**(4), 297–323.

Velásquez, J. D. (1998), Modeling emotion-based decision making, *in* D. Cañamero, ed., 'Emotional and Intelligent: The Tangled Knot of Cognition', AAAI Press.

Vytal, K. & Hamann, S. (2010), 'Neuroimaging support for discrete neural correlates of basic emotions: A voxel-based meta-analysis', *Journal of Cognitive Neuroscience* **22**(12), 2864–2885.

Wallis, J. D. (2007), 'Orbitofrontal cortex and its contribution to decision-making', *Annual Review of Neuroscience* **30**, 31–56.

Watson, D., Clark, L. A. & Tellegen, A. (1988), 'Development and validation of brief measures of positive and negative affect: The panas scales', *Journal of Personality and Social Psychology* **54**(6), 1063–1070.

Wicker, F. W., Brown, G., Wiehe, J. A., Hagen, A. S. & Reed, J. L. (1993), 'On reconsidering Maslow: An examination of the deprivation/dominance proposition', *Journal of Research in Personality* **27**(2), 118–133.

Zajonc, R. B. (1984), 'Feeling and thinking: Preferences need no inferences', *American Psychologist* **35**(2), 151–175.

Appendix A

Vytal's review of neurological discrete emotion studies

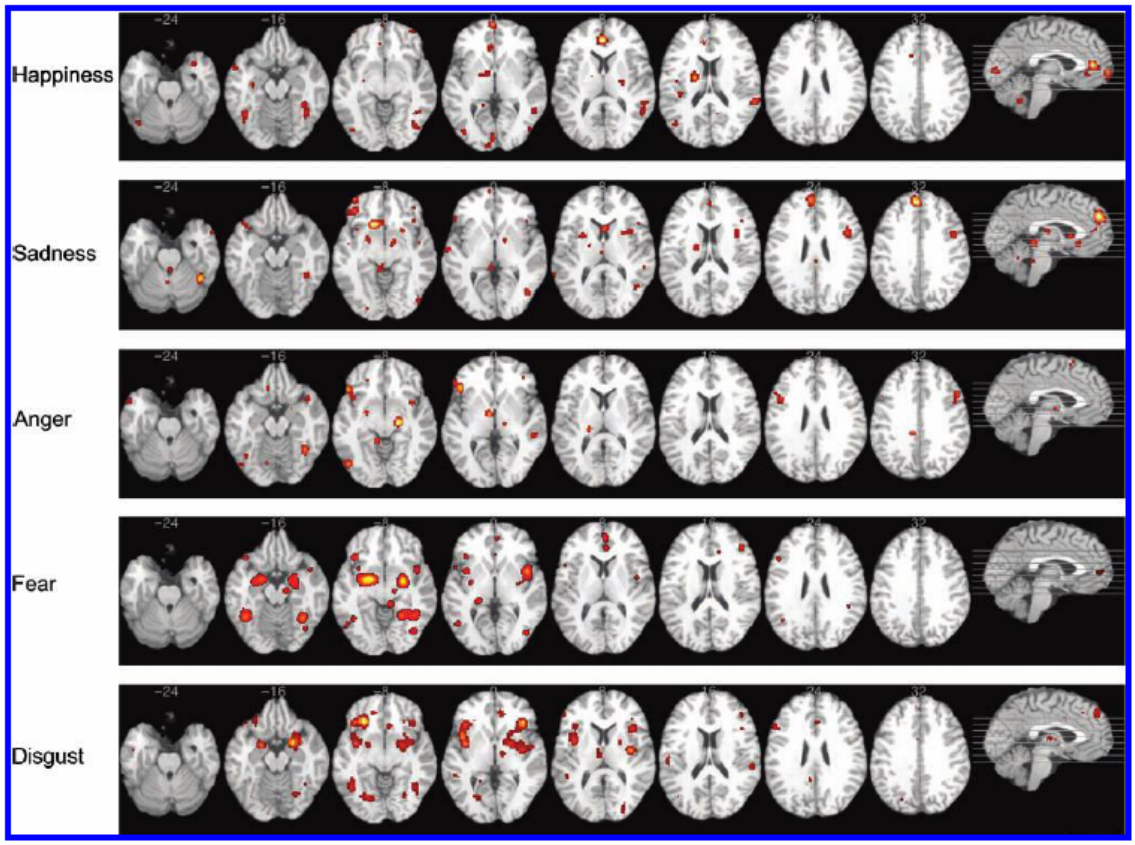


Figure A.1: Brain activation likelihood maps for discrete emotions (Vytal & Hamann 2010)

Appendix B

Ortony's list of basic emotion studies

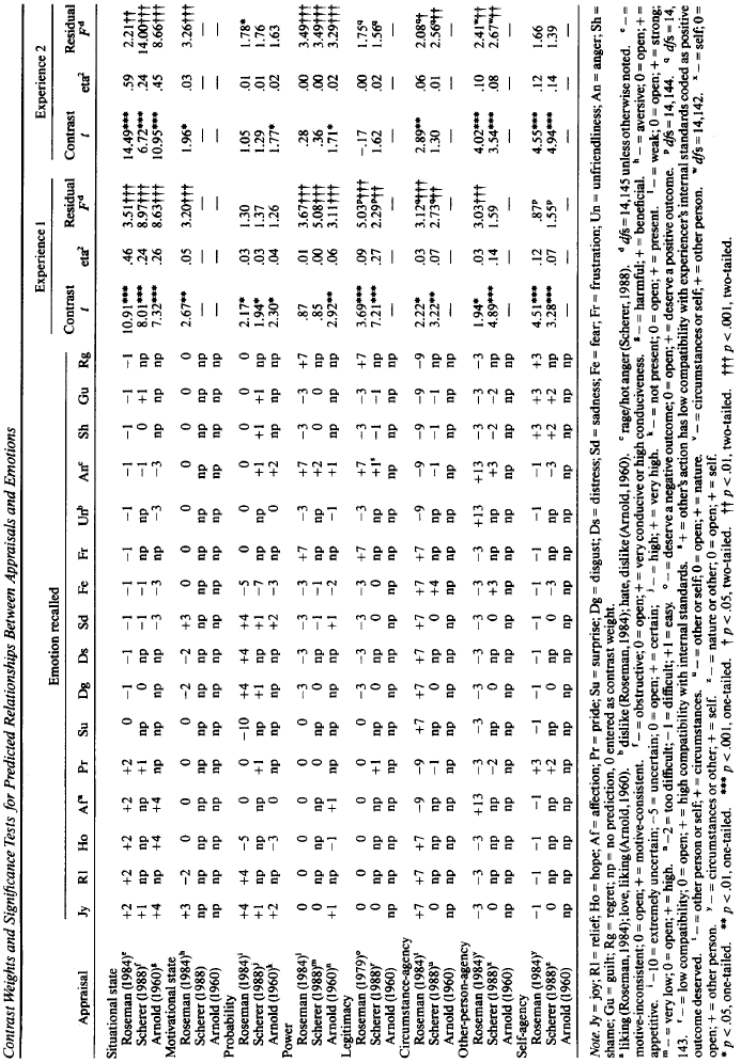
A Selection of Lists of "Basic" Emotions

Reference	Fundamental emotion	Basis for inclusion
Arnold (1960)	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness	Relation to action tendencies
Ekman, Friesen, & Ellsworth (1982)	Anger, disgust, fear, joy, sadness, surprise	Universal facial expressions
Frijda (personal communication, September 8, 1986)	Desire, happiness, interest, surprise, wonder, sorrow	Forms of action readiness
Gray (1982)	Rage and terror, anxiety, joy	Hardwired
Izard (1971)	Anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise	Hardwired
James (1884)	Fear, grief, love, rage	Bodily involvement
McDougall (1926)	Anger, disgust, elation, fear, subjection, tender-emotion, wonder	Relation to instincts
Mowrer (1960)	Pain, pleasure	Unlearned emotional states
Oatley & Johnson-Laird (1987)	Anger, disgust, anxiety, happiness, sadness	Do not require propositional content
Panksepp (1982)	Expectancy, fear, rage, panic	Hardwired
Plutchik (1980)	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise	Relation to adaptive biological processes
Tomkins (1984)	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise	Density of neural firing
Watson (1930)	Fear, love, rage	Hardwired
Weiner & Graham (1984)	Happiness, sadness	Attribution independent

Figure B.1: Different studies of basic emotions (Ortony & Turner 1990)

Appendix C

Roseman's event appraisal experiment results



Contrast Weights and Significance Tests for Predicted Relationships Between Appraisals and Emotions

Appraisal	Emotion recalled															Experience 1			Experience 2		
	Jy	Rl	Ho	Af[a]	Pr	Su	Dg	Ds	Sd	Fe	Fr	Ufr[b]	Auf[c]	Sh	Gu	Rg	Contrast *t*	eta[2]	Residual F^{d}	Contrast *t*	eta[2]	Residual F^{d}
Situational state																						
Roseman (1984)[e]	+2	+2	+2	+2	+2	0	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	10.91***	.46	3.51†††	14.49***	.59	2.21††
Scherer (1988)[f]	+1	np	np	np	+1	np	0	np	−1	−1	np	np	−1	0	+1	np	8.01***	.24	8.97†††	6.72***	.24	14.00†††
Arnold (1960)[g]	+4	np	+4	np	np	np	np	−3	−3	np	np	−3	np	np	np	np	7.32***	.26	8.63†††	10.95***	.45	8.66†††
Motivational state																						
Roseman (1984)[h]	+3	−2	0	0	0	0	−2	−2	+3	0	0	0	0	0	0	0	2.67**	.05	3.20†††	1.96*	.03	3.26†††
Scherer (1988)	np	np	np	np	np	np	np	np	np	np	np	np	np	np	np	np	—	—	—	—	—	—
Arnold (1960)	np	np	np	np	np	np	np	np	np	np	np	np	np	np	np	np	—	—	—	—	—	—
Probability																						
Roseman (1984)[i]	+4	+4	−5	0	−10	+4	+4	+4	−5	0	0	0	0	0	0	0	2.17*	.03	1.30	1.05	.01	1.78*
Scherer (1988)[j]	+1	np	np	+1	np	+1	np	+1	−7	np	np	np	+1	+1	+1	np	1.94*	.03	1.37	1.29	.01	1.76
Arnold (1960)[k]	+2	np	−3	0	np	np	np	+2	−3	np	0	np	+2	np	np	np	2.30*	.04	1.26	1.77*	.02	1.63
Power																						
Roseman (1984)[l]	0	0	0	0	0	0	−3	−3	−3	−3	+7	−3	+7	−3	−3	+7	.87	.01	3.67†††	.28	.00	3.49†††
Scherer (1988)[m]	0	np	np	np	0	np	0	np	−1	−1	np	np	+2	0	0	np	.85	.00	5.08†††	.36	.00	3.49†††
Arnold (1960)[n]	+1	np	−1	+1	np	np	np	np	+1	−2	np	−1	+1	np	np	np	2.92**	.06	3.11†††	1.71*	.02	3.29†††
Legitimacy																						
Roseman (1979)[o]	0	0	0	0	0	0	−3	−3	−3	−3	+7	−3	+7	−3	−3	+7	3.69***	.09	5.03†††	−.17	.00	1.75[s]
Scherer (1988)[f]	0	np	np	np	+1	np	0	np	0	0	np	np	+1[r]	−1	−1	np	7.21***	.27	2.29††	1.62	.02	1.56[s]
Arnold (1960)	np	np	np	np	np	np	np	np	np	np	np	np	np	np	np	np	—	—	—	—	—	—
Circumstance-agency																						
Roseman (1984)[e]	+7	+7	+7	−9	−9	+7	+7	+7	+7	+7	+7	−9	−9	−9	−9	−9	2.22*	.03	3.12†††	2.89**	.06	2.08††
Scherer (1988)[f]	0	np	np	np	−1	np	0	np	0	+4	np	np	−1	−1	−1	np	3.22**	.07	2.73††	1.30	.01	2.56††
Arnold (1960)	np	np	np	np	np	np	np	np	np	np	np	np	np	np	np	np	—	—	—	—	—	—
Other-person-agency																						
Roseman (1984)[e]	−3	−3	−3	+13	−3	−3	−3	−3	−3	−3	−3	+13	+13	−3	−3	−3	1.94*	.03	3.03†††	4.02***	.10	2.41†††
Scherer (1988)[f]	0	np	np	np	−2	np	0	np	0	+3	np	np	+3	−2	−2	np	4.89***	.14	1.59	3.54***	.08	2.67†††
Arnold (1960)	np	np	np	np	np	np	np	np	np	np	np	np	np	np	np	np	—	—	—	—	—	—
Self-agency																						
Roseman (1984)[e]	−1	−1	−1	−1	+3	−1	−1	−1	−1	−1	−1	−1	−1	+3	+3	+3	4.51***	.12	.87[p]	4.55***	.12	1.66
Scherer (1988)[f]	0	np	np	np	+2	np	0	np	0	−3	np	np	−3	+2	+2	np	3.28***	.07	1.55[p]	4.94***	.14	1.39
Arnold (1960)	np	np	np	np	np	np	np	np	np	np	np	np	np	np	np	np	—	—	—	—	—	—

Note. Jy = joy; Rl = relief; Ho = hope; Af = affection; Pr = pride; Su = surprise; Dg = disgust; Ds = distress; Sd = sadness; Fe = fear; Fr = frustration; Un = unfriendliness; An = anger; Sh = shame; Gu = guilt; Rg = regret; np = no prediction, 0 entered as contrast weight.

[a] liking (Roseman, 1984); love, liking (Arnold, 1960). [b] dislike (Roseman, 1984); hate, dislike (Arnold, 1960). [c] rage/hot anger (Scherer, 1988). [d] dfs = 14, 145 unless otherwise noted. [e] − = motive-inconsistent; 0 = open; + = motive-consistent. [f] − = obstructive; 0 = open; + = very conducive or high conduciveness. [g] − = harmful; + = beneficial. [h] − = aversive; 0 = open; + = appetitive. [i] −10 = extremely uncertain; −5 = uncertain; 0 = open; + = certain; [j] − = high; + = very high. [k] − = not present; 0 = open; + = present. [l] − = weak; 0 = open; + = strong; [m] − = very low; 0 = open; + = high. [n] −2 = too difficult; −1 = difficult; +1 = easy. [o] − = deserve a negative outcome; 0 = open; + = deserve a positive outcome. [p] dfs = 14, 144. [q] dfs = 14, 143. [r] − = low compatibility; 0 = open; + = high compatibility with internal standards. [s] + = other's action has low compatibility with experiencer's internal standards coded as positive outcome deserved. [t] − = other person or self; 0 = open; + = nature. [u] − = circumstances or self; + = other person. [v] dfs = 14, 142. [w] − = self; 0 = open; + = other person. [x] − = circumstances or other; + = self. [y] − = nature or other; 0 = open; + = self.

[z] $p < .05$, one-tailed. ** $p < .01$, one-tailed. *** $p < .001$, one-tailed. † $p < .05$, two-tailed. †† $p < .01$, two-tailed. ††† $p < .001$, two-tailed.

Figure C.1: Mean appraisal ratings and significance tests for appraisal dimensions (Roseman et al. 1990)

Appendix D

The model usage diagram

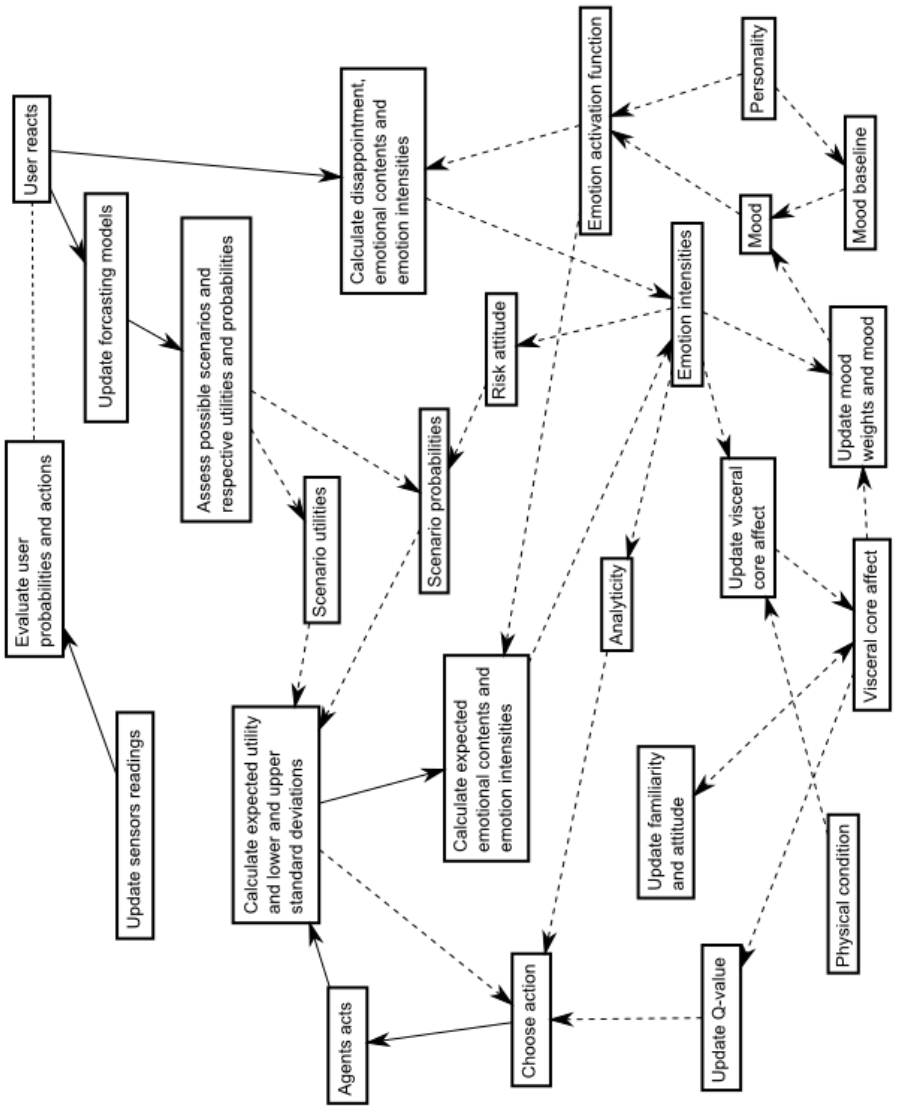


Figure D.1: The model usage process with solid arrows indicating progress and dashed arrows indicating data flow